

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 19:05:52 ; Search time 4068 Seconds  
(without alignments)  
10444.960 Million cell updates/sec

Title: US-09-595-947C-1  
Perfect score: 1460  
Sequence: 1 gcaggtacgagagagcag.....agagtacctaataccagtg 1460

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2034640 seqs, 14551402878 residues

Word size : 15

Total number of hits satisfying chosen parameters: 42842

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_nam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1460	100.0	1491	6	A91167	A91167 Sequence 1
2	1460	100.0	1491	10	RNRELAXT	Y10619 R.norvegicus
3	634	43.4	166677	2	AC127817	AC127817 Rattus no
4	92	6.3	861	10	MMU76208	U76208 Mus musculus
5	92	6.3	1861	10	AF364300	AF364300 Mus muscu
6	92	6.3	5567	10	MMATH4B	Y0167 M.musculus
7	92	6.3	138070	2	AC109783	AC109783 Mus muscu
8	92	6.3	185806	2	AC127417	AC127417 Mus muscu
9	35	2.4	6123	9	AF303002	AF303002 Homo sapi
10	35	2.4	91531	2	AC079846_3	Continuation (4 of
11	35	2.4	179697	9	AC023886	AC023886 Homo sapi
12	32	2.2	1330	9	HSAL33776	AL33776 Homo sapi
13	32	2.2	5340	9	AF234829	AF234829 Homo sapi
14	32	2.2	165110	9	AL450311	AL450311 Human DNA
15	32	2.2	173341	2	AC021954	AC021954 Homo sapi
16	26	1.8	170896	2	AC011010	AC011010 Homo sapi
17	25	1.7	25	6	A91170	A91170 Sequence 4
18	23	1.6	735	10	MMU67776	U67776 Mus musculus
19	23	1.6	770	5	AF123884	AF123884 Gallus ga
20	23	1.6	790	5	GGA012659	AJ012659 Gallus ga
21	23	1.6	932	10	MMATH4C	Y09166 M.musculus
22	23	1.6	1315	10	MMU63841	U63841 Mus musculus
23	23	1.6	1333	6	AR023715	AR023715 Sequence
24	23	1.6	1341	5	AF109014	AF109014 Gallus ga
25	23	1.6	1385	10	MMU76207	U76207 Mus musculus
26	23	1.6	1412	10	MMATH4A	Y07621 M.musculus
27	23	1.6	1880	5	AF303000	AF303000 Gallus ga
28	23	1.6	10393	10	AF303001	AF303001 Mus muscu
29	23	1.6	71538	2	AC118243	AC118243 Mus muscu
30	23	1.6	123855	2	AC102600	AC102600 Mus muscu
31	23	1.6	174688	2	AC124395	AC124395 Mus muscu
32	21	1.4	310	6	I84656	I84656 Sequence 12
33	21	1.4	1164	5	GGNEURODL	Y09596 G.gallus mR
34	21	1.4	1268	6	AR023709	AR023709 Sequence
35	21	1.4	1268	9	HSU63842	U63842 Human neuro
36	21	1.4	1352	6	I84655	I84655 Sequence 10
37	21	1.4	1535	6	AR023708	AR023708 Sequence
38	21	1.4	1550	6	AR103242	AR103242 Sequence
39	21	1.4	1550	6	BD000105	BD000105 Different
40	21	1.4	1550	9	HSU58681	U58681 Homo sapien
41	21	1.4	1596	5	AF060885	AF060885 Gallus ga
42	21	1.4	1675	9	BC008687	BC008687 Homo sapi
43	21	1.4	1717	9	BC028226	BC028226 Homo sapi
44	21	1.4	1831	10	MUSNDRF	D83507 Mouse mRNa
45	21	1.4	1843	10	RAT4	D82868 Rattus norv

ALIGNMENTS

RESULT 1	A91167	Sequence 1 from Patent WO9827206.	1491 bp	DNA	linear	PAT 22-JAN-2000
LOCUS	A91167	Sequence 1 from Patent WO9827206.	1491 bp	DNA	linear	PAT 22-JAN-2000
DEFINITION	A91167	Sequence 1 from Patent WO9827206.	1491 bp	DNA	linear	PAT 22-JAN-2000
ACCESSION	A91167	Sequence 1 from Patent WO9827206.	1491 bp	DNA	linear	PAT 22-JAN-2000
VERSION	A91167.1	GI:6740202	1491 bp	DNA	linear	PAT 22-JAN-2000
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
JOURNAL						

BEST AVAILABLE COPY

ICARD LIEPKALNS CHRISTINE (FR); MALLET JACQUES (FR)

## FEATURES

Location/Qualifiers

1. .1491

/organism="Rattus sp."

/db\_xref="taxon:10118"

459. .1103

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAP69414.1"

/db\_xref="GI:6740203"

DSEARAGCRGTSKLRARRGRNRPKSELALSKORRRKANDRENRHNLNSA

LQALRGVLPFDKLTIELRLFAHNTLWTLRIADHSFYGPPEPVPCEIGLS

PGGSGDNGWSIYSPVQAGSLSPSTALEFFGLQVPSPSCLLPGLTVFSDFL"

BASE COUNT 307 a 487 c 413 g 284 t

## ORIGIN

Query Match 100.0%; Score 1460; DB 6; Length 1491;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGTAGCGAGAGGACGCTCTGGCCGCCCGTTCCTGATTTGGCCCGTGGCACAGGCA 60  
 DB 1 GCAGGTAGCGAGAGGACGCTCTGGCCGCCCGTTCCTGATTTGGCCCGTGGCACAGGCA 60

QY 61 GCAGCCGGCAGGACGCTCTGGTCCGGCAGACAGATAAGCGTGCAGGGGACACA 120  
 DB 61 GCAGCCGGCAGGACGCTCTGGTCCGGCAGACAGATAAGCGTGCAGGGGACACA 120

QY 121 CGATTAGCAGCTCAGAGTCCCTCTGGTCTCACCAGTGCACAGAGCCGAGACCCCT 180  
 DB 121 CGATTAGCAGCTCAGAGTCCCTCTGGTCTCACCAGTGCACAGAGCCGAGACCCCT 180

QY 181 CCGAGCTTCTTTGCTCCAGACGCAATTTACTCCAGGCGAGGGCGCTTCGAGCTCAG 240  
 DB 181 CCGAGCTTCTTTGCTCCAGACGCAATTTACTCCAGGCGAGGGCGCTTCGAGCTCAG 240

QY 241 CAACACTTCGAGGACGAGGGTTACGATTCACCGCTGCTTACGTCGACACCC 300  
 DB 241 CAACACTTCGAGGACGAGGGTTACGATTCACCGCTGCTTACGTCGACACCC 300

QY 301 GCAGCTCTGTCTTTGAGCCGGAGTAAGTAGTAACATTTAGGAACCTCCAAAGGG 360  
 DB 301 GCAGCTCTGTCTTTGAGCCGGAGTAAGTAGTAACATTTAGGAACCTCCAAAGGG 360

QY 361 TAGAAGAGGAGTGGTGGCGTACTCTAGTCCCGGCTGGAGTGACCTCTAAGTCAGAG 420  
 DB 361 TAGAAGAGGAGTGGTGGCGTACTCTAGTCCCGGCTGGAGTGACCTCTAAGTCAGAG 420

QY 421 ACTGTACACCCCTTCCATTTTCCCAACCTCAGAGTGGCGCCCTCATCCCTTGGATG 480  
 DB 421 ACTGTACACCCCTTCCATTTTCCCAACCTCAGAGTGGCGCCCTCATCCCTTGGATG 480

QY 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGCCTCGGACACG 540  
 DB 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGCCTCGGACACG 540

QY 541 AAGTGCTCAGTTCOAATTCACCCACCTAGCCCACTCTCGTACCGAGGAGTGTCTCCG 600  
 DB 541 AAGTGCTCAGTTCOAATTCACCCACCTAGCCCACTCTCGTACCGAGGAGTGTCTCCG 600

QY 601 AAGCAGAGCAGTGTACTCGGAGGACATCGAGGAGTCCGTCGCGCGCGGAGGGC 660  
 DB 601 AAGCAGAGCAGTGTACTCGGAGGACATCGAGGAGTCCGTCGCGCGCGGAGGGC 660

QY 661 GCAACAGGCCCAAGAGCGAGTGTGCTAGCAGAGCAGCAGCAAGCGCGCAAGAGG 720  
 DB 661 GCAACAGGCCCAAGAGCGAGTGTGCTAGCAGAGCAGCAGCAAGCGCGCGCAAGAGG 720

QY 721 CCAACAGCCGGAGCGCAACCGCATGTCACACCTTAACCTCCGGGCTGGATCGCTGCGG 780  
 DB 721 CCAACAGCCGGAGCGCAACCGCATGTCACACCTTAACCTCCGGGCTGGATCGCTGCGG 780

QY 781 GTGTCGTGCCCACTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCTGGCGTTCCG 840  
 DB 781 GTGTCGTGCCCACTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCTGGCGTTCCG 840

QY 841 CCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACACAGCTTCTACG 900  
 DB 841 CCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACACAGCTTCTACG 900

QY 901 GCCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAAGCCCGGAGGGGGTCCACAGCGCG 960  
 DB 901 GCCCCGAGCCCTGTGCCCTGTGGGAGCTGGGAAGCCCGGAGGGGGTCCACAGCGCG 960

QY 961 ACTGGGGCTCTATCTACTCTCCAGTTTCCCAAGCTGTAGCCTGAGCCCAACAGCTCAT 1020  
 DB 961 ACTGGGGCTCTATCTACTCTCCAGTTTCCCAAGCTGTAGCCTGAGCCCAACAGCTCAT 1020

QY 1021 TGGAGGAGTTCCCTGGCTGCAGGTGCCAGCTCCCATCTCTGTCCTCCCGGGCACCC 1080  
 DB 1021 TGGAGGAGTTCCCTGGCTGCAGGTGCCAGCTCCCATCTCTGTCCTCCCGGGCACCC 1080

QY 1081 TGTGTCTCTCAGACTTCTTGTGAAGGCCCCAAACAGAGCCCTGGCGGTGGCGCTGGCAG 1140  
 DB 1081 TGTGTCTCTCAGACTTCTTGTGAAGGCCCCAAACAGAGCCCTGGCGGTGGCGCTGGCAG 1140

QY 1141 AAAGGAGGAGTGCAGAGCTGTCTGAAATGGAAGTGTAGTGAGGCACCTCGAGCATCTCGC 1200  
 DB 1141 AAAGGAGGAGTGCAGAGCTGTCTGAAATGGAAGTGTAGTGAGGCACCTCGAGCATCTCGC 1200

QY 1201 CCTTCTGGCTTTTCATTAGTCAGTCCCTGATTTAAACAGGATTTCGCACAGTTCCTTGGCT 1260  
 DB 1201 CCTTCTGGCTTTTCATTAGTCAGTCCCTGATTTAAACAGGATTTCGCACAGTTCCTTGGCT 1260

QY 1261 GCTGTGCGTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTTCCTCAGTGTGGCC 1320  
 DB 1261 GCTGTGCGTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTTCCTCAGTGTGGCC 1320

QY 1321 ACCTCAACTCCCGCTCCAGCAGAGGAGAGCCGTACACATAATAGTTGGAGACTCCC 1380  
 DB 1321 ACCTCAACTCCCGCTCCAGCAGAGGAGAGCCGTACACATAATAGTTGGAGACTCCC 1380

QY 1381 ATACTTCTGGTGTACTCCGCCCTCTTTCAAATCTGCGGGCTCCCAACACCGCTTTCTCC 1440  
 DB 1381 ATACTTCTGGTGTACTCCGCCCTCTTTCAAATCTGCGGGCTCCCAACACCGCTTTCTCC 1440

QY 1441 AGAGTGACCTTAATCCAGTGT 1460  
 DB 1441 AGAGTGACCTTAATCCAGTGT 1460

## RESULT 2

## RNRELAXT

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

RNRELAXT 1491 bp mRNA linear ROD 06-MAY-1997  
 R.norvegicus mRNA for transcriptional regulator, Relax.

Y10619  
 Y10619.1 GI:2072737

Relax; transcriptional regulator.

Rattus norvegicus.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 1491)

Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.

Relax, a novel rat bHLH transcriptional regulator transiently

expressed in the ventricular proliferating zone of the developing

central nervous system

J. Neurosci. Res. 48 (2), 146-158 (1997)

97276390

9130143

2 (bases 1 to 1491)

Ravassard, P.

Direct Submission

Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,



Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Correll, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lied, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
 Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 166677)  
 Worley, K.C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GZXS  
 Center clone name: CH230-259c16  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 73194 bases at least Q40  
 Consensus quality: 79767 bases at least Q30  
 Consensus quality: 84278 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 80 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1032: contig of 1032 bp in length  
 \* 1033 1132: gap of unknown length  
 \* 1133 2146: contig of 1014 bp in length  
 \* 2147 2246: gap of unknown length  
 \* 2247 3549: contig of 1303 bp in length  
 \* 3550 3649: gap of unknown length  
 \* 3650 4695: contig of 1046 bp in length  
 \* 4696 4795: gap of unknown length  
 \* 4796 4896: contig of 1397 bp in length  
 \* 4897 6192: gap of unknown length  
 \* 6193 6292: contig of 1675 bp in length  
 \* 6293 7967: gap of unknown length  
 \* 7968 8067: gap of unknown length  
 \* 8068 9286: contig of 1219 bp in length  
 \* 9287 9386: gap of unknown length  
 \* 9387 10392: contig of 1006 bp in length  
 \* 10393 10492: gap of unknown length  
 \* 10493 11602: contig of 1110 bp in length  
 \* 11603 11702: gap of unknown length  
 \* 11703 13173: contig of 1471 bp in length  
 \* 13174 13273: gap of unknown length  
 \* 13274 14294: contig of 1021 bp in length  
 \* 14295 14394: gap of unknown length  
 \* 14395 15982: contig of 1588 bp in length  
 \* 15983 16082: gap of unknown length  
 \* 16083 17115: contig of 1033 bp in length  
 \* 17116 17215: gap of unknown length  
 \* 17216 18361: contig of 1146 bp in length  
 \* 18362 18461: gap of unknown length  
 \* 18462 20372: contig of 1910 bp in length  
 \* 20373 21519: contig of 1048 bp in length  
 \* 21520 21619: gap of unknown length  
 \* 21620 22638: contig of 1019 bp in length  
 \* 22639 22738: gap of unknown length  
 \* 22739 23988: contig of 1250 bp in length  
 \* 23989 24088: gap of unknown length  
 \* 24089 25265: contig of 1177 bp in length  
 \* 25266 25365: gap of unknown length  
 \* 25366 26675: contig of 1310 bp in length  
 \* 26676 26775: gap of unknown length  
 \* 26776 28414: contig of 1639 bp in length  
 \* 28415 28514: gap of unknown length  
 \* 28515 29921: contig of 1407 bp in length  
 \* 29922 30021: gap of unknown length  
 \* 30022 31951: contig of 1930 bp in length  
 \* 31952 32051: gap of unknown length  
 \* 32052 33980: contig of 1929 bp in length  
 \* 33981 34080: gap of unknown length  
 \* 34081 35824: contig of 1744 bp in length  
 \* 35825 35924: gap of unknown length  
 \* 35925 37572: contig of 1648 bp in length  
 \* 37573 37672: gap of unknown length  
 \* 37673 39232: contig of 1560 bp in length  
 \* 39233 39332: gap of unknown length  
 \* 39333 41121: contig of 1788 bp in length  
 \* 41122 42662: contig of 1442 bp in length  
 \* 42663 42762: gap of unknown length  
 \* 42763 44647: contig of 1885 bp in length  
 \* 44648 46211: contig of 1464 bp in length  
 \* 46212 46311: gap of unknown length  
 \* 46312 48275: contig of 1964 bp in length  
 \* 48276 48375: gap of unknown length  
 \* 48376 49450: contig of 1075 bp in length  
 \* 49451 49550: gap of unknown length  
 \* 49551 51299: contig of 1749 bp in length  
 \* 51300 51399: gap of unknown length  
 \* 51400 52749: contig of 1350 bp in length  
 \* 52750 52849: gap of unknown length  
 \* 52850 54511: contig of 1662 bp in length  
 \* 54512 54611: gap of unknown length  
 \* 54612 55928: contig of 1317 bp in length  
 \* 55929 56028: gap of unknown length  
 \* 56029 57531: contig of 1503 bp in length  
 \* 57532 57631: gap of unknown length  
 \* 57632 59080: contig of 1449 bp in length



```

* 59081 59180: gap of unknown length
* 59181 60513: contig of 1333 bp in length
* 60514 60613: gap of unknown length
* 60614 62621: contig of 2008 bp in length
* 62622 62721: gap of unknown length
* 62722 64707: contig of 1986 bp in length
* 64708 64807: gap of unknown length
* 64808 66176: contig of 1369 bp in length
* 66177 66276: gap of unknown length
* 66277 68684: contig of 2408 bp in length
* 68685 68784: gap of unknown length
* 68785 70841: contig of 2057 bp in length
* 70842 70942: gap of unknown length
* 70942 72278: contig of 1337 bp in length
* 72279 72379: gap of unknown length
* 72379 73668: contig of 1490 bp in length
* 73669 73969: gap of unknown length
* 73969 75979: contig of 2011 bp in length
* 75980 76079: gap of unknown length
* 76080 78068: contig of 1989 bp in length
* 78069 78168: gap of unknown length
* 78169 79454: contig of 1286 bp in length
* 79455 79554: gap of unknown length
* 79555 80709: contig of 1155 bp in length
* 80710 80809: gap of unknown length
* 80810 82754: contig of 1945 bp in length
* 82755 82854: gap of unknown length
* 82855 83396: contig of 2542 bp in length
* 83397 85496: gap of unknown length
* 85497 88065: contig of 2569 bp in length
* 88066 88165: gap of unknown length
* 88166 90996: contig of 2831 bp in length
* 90997 91096: gap of unknown length
* 91097 92377: contig of 1481 bp in length
* 92378 92678: gap of unknown length
* 92678 94520: contig of 1843 bp in length
* 94521 94620: gap of unknown length

Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 GACCTCGCCTTGCCGCCACAACACTACATTTGGGCACCTGACTCAGACGCTGCGCATAGCGGA 886
DB 57641 GACCTCGCCTTGCCGCCACAACACTACATTTGGGCACCTGACTCAGACGCTGCGCATAGCGGA 57700

QY 887 CCACAGCTTCTACGGCCCGCGAGCCCGCTGCGCTGTGGGAGCTGGGAAGCCCGGGAGG 946
DB 57701 CCACAGCTTCTACGGCCCGCGAGCCCGCTGCGCTGTGGGAGCTGGGAAGCCCGGGAGG 57760

QY 947 GGGCTCCAGCGGACTGGGGCTCTATCTACTCCCGAGTTTCCCAAGCTGGTAGCCTGAG 1006
DB 57761 GGGCTCCAGCGGACTGGGGCTCTATCTACTCCCGAGTTTCCCAAGCTGGTAGCCTGAG 57820

QY 1007 CCCACAGCCTCATTTGAGGAGTTCCTGCGCTGCGCTGAGTCCCGAGCTCCCGATCCTGTCT 1066
DB 57821 CCCACAGCCTCATTTGAGGAGTTCCTGCGCTGCGCTGAGTCCCGAGCTCCCGATCCTGTCT 57880

QY 1067 GCTCCGGGACCCCTGGTCTTCTCAGACTCTTGTGAAGGGCCCAACAGCGCCCTGGGG 1126
DB 57881 GCTCCGGGACCCCTGGTCTTCTCAGACTCTTGTGAAGGGCCCAACAGCGCCCTGGGG 57940

QY 1127 GTGGCGCTGGCAGAAAGGAGGAGTCAAGCTGCTCAAAATGGAAGTGTGGAGGCA 1186
DB 57941 GTGGCGCTGGCAGAAAGGAGGAGTCAAGCTGCTCAAAATGGAAGTGTGGAGGCA 58000

QY 1187 CTCGAGCATCTCGCCCTCTCTGGCTTTTATTAGTCAGGTCCTGATTTAACCAAGATTGG 1246
DB 58001 CTCGAGCATCTCGCCCTCTCTGGCTTTTATTAGTCAGGTCCTGATTTAACCAAGATTGG 58060

QY 1247 CACAGTTCCTGCTGTGTGCTGCACAAAGGACATTGACAGCTGATCTCTTTAAACC 1306
DB 58061 CACAGTTCCTGCTGTGTGCTGCACAAAGGACATTGACAGCTGATCTCTTTAAACC 58120

```

```

QY 1307 TCCTCAGTGTGGCCACTCAAACTCCCGCTCCCAAGCAGAGAGAGCGTAGCCTAAATA 1366
DB 58121 TCCTCAGTGTGGCCACTCAAACTCCCGCTCCCAAGCAGAGAGAGCGTAGCCTAAATA 58180

QY 1367 GTTGGGAGACTCCCATCTTCCTGGTGACTCCGCCCTCTTTCAAATCTCGGGCCTCAA 1426
DB 58181 GTTGGGAGACTCCCATCTTCCTGGTGACTCCGCCCTCTTTCAAATCTCGGGCCTCAA 58240

QY 1427 CCACCGCTTTCTCCAGAGTGACCTTAATCCAGTGT 1460
DB 58241 CCACCGCTTTCTCCAGAGTGACCTTAATCCAGTGT 58274

RESULT 4
MMU76208 861 bp DNA linear ROD 05-FEB-1997
LOCUS Mus musculus neurogenin 3 (ngn3) gene, complete cds.
DEFINITION
ACCESSION U76208
VERSION U76208.1 GI:1815654
KEYWORDS
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 861)
AUTHORS Sommer,L., Ma,Q. and Anderson,D.J.
TITLE Neurogenins, a novel family of atonal-related bHLH transcription
factors, are putative mammalian neuronal determination genes that
reveal progenitor cell heterogeneity in the developing CNS and PNS
JOURNAL Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
MEDLINE 97153585
PUBMED 9000438
REFERENCE 2 (bases 1 to 861)
AUTHORS Sommer,L., Ma,Q. and Anderson,D.J.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1996) Biology 216-76, California Institute of
Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
USA
COMMENT On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES
source
1..861
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 150..804
/gene="ngn3"
CDS 150..804
/function="bHLH transcription factor"
/note="NGN3"
/codon_start=1
/product="neurogenin 3"
/protein_id="AAC53029.1"
/db_xref="GI:1666912"
/translation="MAPHPLDALTIQVSPETQOPFGASDHEVLSSNTPSPPTLIPR
DCSEAVGDCRGTSRKLRRGRNRKPSKSELALSKORRRKKKANDRRNRHNLNSA
LDALRGVLPDPDDAKLKIETLFAHNV IWA L T O T L R I A D H S F Y G P P V C G E L G S
P G G S N G D W G S I Y S P V S G A G N L S P A S L E E F F G L Q V P S P S Y L L P G A L V F S D F L "
BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATCGCTCGCGGTGTCTCTGCCACCTTCCCGATGACGCCAACCTTACAAAG 821
DB 463 GCGCTGGATCGCTCGCGGTGTCTCTGCCACCTTCCCGATGACGCCAACCTTACAAAG 522

QY 822 ATCGAGACCCCTCGGCTTCGCCCCACAACACTACAT 853
DB 523 ATCGAGACCCCTCGGCTTCGCCCCACAACACTACAT 554

```

```

RESULT 5
AF364300
LOCUS
DEFINITION Mus musculus neurogenin 3 gene, complete cds.
ACCESSION AF364300
VERSION AF364300.1 GI:13937128
KEYWORDS
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1861)
AUTHORS Lee, J., Smith, S., Watada, H., Lin, J., Scheel, D., Wang, J., Mirmira, R.
and German, M.
TITLE Regulation of the pancreatic pro-endocrine gene neurogenin3
JOURNAL Diabetes (2001) In press
REFERENCE 2 (bases 1 to 1861)
AUTHORS Schwitzgebel, V. and German, M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2001) Hormone Research Institute, University of
California San Francisco, 513 Parnassus Ave, HSW1090, San
Francisco, CA 94145-0534, USA
FEATURES
source
1. .1861
/organism="Mus musculus"
/db_xref="taxon:10090"
<1093..>1737
/product="neurogenin 3"
1093..1737
/codon_start=1
/product="neurogenin 3"
/protein_id="AA850058.1"
/db_xref="GI:13937129"
/translocation="MAPHPLDALTIOVSPETQPPFGASDHEVLSSNTPPSTLPIR
DCSEAEVDCRGTSRKLRARRGNRPKSELALSQRSSRRKKANDRRNRMNLSA
LDALRGVLPFPDDAKLTETLRFAHNYIWTOTLRADHSFYGPPEVPVCGELGS
PGGSGNDWGSISYSPYSQAGNLSPASLEEFPGVPSYLLPGALVFSDFL"
BASE COUNT 397 a 560 c 537 g 367 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 1861;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821
|||||
Db 1396 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 1455
|||||

QY 822 ATCGAGACCTTGGCTTCGCCCACTACAT 853
|||||
Db 1456 ATCGAGACCTTGGCTTCGCCCACTACAT 1487
|||||

RESULT 6
MMATH4B
LOCUS
DEFINITION M musculus MATH4B gene.
ACCESSION Y09167
VERSION Y09167.2 GI:11065737
KEYWORDS bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 5567)
AUTHORS Cau, E., Gradwohl, G., Fode, C. and Guillemot, F.
TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron
progenitors
JOURNAL Development 124 (8), 1611-1621 (1997)
MEDLINE 97261963
PUBMED 9108377
REFERENCE 2 (bases 1 to 5567)

```

```

AUTHORS Jacquemin, P., Durviaux, S.M., Jensen, J., Godfraind, C., Gradwohl, G.,
Guillemot, F., Madsen, O.D., Carmeliet, P., Dewerchin, M., Collen, D.,
Rousseau, G. and Lemaigre, F.P.
TITLE Transcription factor hepatocyte nuclear factor 6 regulates
pancreatic endocrine cell differentiation and controls expression
of the proendocrine gene ngn3
JOURNAL Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
MEDLINE 20285449
PUBMED 10825208
REFERENCE 3 (bases 1 to 5567)
AUTHORS Gradwohl, G.J.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,
F-67404 ILLKIRCH cedex, FRANCE
REMARK Revised by [4]
REFERENCE 4 (bases 1 to 5567)
AUTHORS Lemaigre, F.P.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
Research Unit, Louvain University Medical School, Avenue Hippocrate
75, box 7529, Brussels 1200, BELGIUM
COMMENT On Oct 31, 2000 this sequence version replaced gi:1666087.
FEATURES
Location/Qualifiers
source
1. .5567
/organism="Mus musculus"
/strain="SV129"
/db_xref="taxon:10090"
/clone="lambda GEM12 clone 19C"
/cell_line="D3"
/cell_type="embryonic stem cells"
/clone_lib="Genomic library No.62, MboI partial digests
cloned in BamHI"
/dev_stage="embryos"
958
misc_binding
582
misc_binding
3692
gene
4145..5567
/gene="MATH4B"
4145..4151
/gene="MATH4B"
/feature="putative"
4923..5567
/gene="MATH4B"
/codon_start=1
/product="neurogenin 3"
/protein_id="CA70366.1"
/db_xref="GI:1666088"
/db_xref="SWISS-PROT:P70661"
/translocation="MAPHPLDALTIOVSPETQPPFGASDHEVLSSNTPPSTLPIR
DCSEAEVDCRGTSRKLRARRGNRPKSELALSQRSSRRKKANDRRNRMNLSA
LDALRGVLPFPDDAKLTETLRFAHNYIWTOTLRADHSFYGPPEVPVCGELGS
PGGSGNDWGSISYSPYSQAGNLSPASLEEFPGVPSYLLPGALVFSDFL"
BASE COUNT 1271 a 1549 c 1564 g 1183 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 5567;
Best Local Similarity 100.0%; Pred. No. 1.4e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821
|||||
Db 5226 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 5285
|||||

QY 822 ATCGAGACCTTGGCTTCGCCCACTACAT 853
|||||
Db 5286 ATCGAGACCTTGGCTTCGCCCACTACAT 5317
|||||

RESULT 7
AC109783/c
LOCUS AC109783 138070 bp DNA linear HTG 07-FEB-2002

```

```

DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
unordered pieces.
AC109783
AC109783.1 GI:18581594
VERSION HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 138070)
McCombie,W.R., Baker,J.P., Ballija,V., Dedhia,N.N., de la
Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
Zutavern,T.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 138070)
McCombie,W.R.
Direct Submission
Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
----- Genome Center -----
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Project name: RP23-121F10
Clone name: RP23-121F10
Insert size: 173000; agarose-fp
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 3.70 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 17785: contig of 17785 bp in length
* 17786 17874: gap of unknown length
* 17875 30356: contig of 12482 bp in length
* 30357 30444: gap of unknown length
* 30445 42306: contig of 11862 bp in length
* 42307 42394: gap of unknown length
* 42395 53598: contig of 11204 bp in length
* 53599 53686: gap of unknown length
* 53687 64355: contig of 10669 bp in length
* 64356 64443: gap of unknown length
* 64444 74017: contig of 9573 bp in length
* 74017 74104: gap of unknown length
* 74105 83366: contig of 9262 bp in length
* 83367 83454: gap of unknown length
* 83455 92355: contig of 8901 bp in length
* 92356 92444: gap of unknown length
* 92444 100821: contig of 8378 bp in length
* 100822 100909: gap of unknown length
* 100910 107529: contig of 6620 bp in length
* 107530 107617: gap of unknown length
* 107618 114066: contig of 6449 bp in length
* 114067 114154: gap of unknown length
* 114155 118873: contig of 4719 bp in length
* 118874 118961: gap of unknown length
* 118962 123619: contig of 4658 bp in length
* 123620 123707: gap of unknown length
* 123708 128240: contig of 4533 bp in length
* 128241 128328: gap of unknown length

```

```

FEATURES
source
1..138070
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-121F10"
BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN
Query Match 6.3% Score 92; DB 2; Length 138070;
Best Local Similarity 100.0%; Pred. No. 1.2e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 762 GCCTGGATCGCTGCGCGTGTCTGCCACCTTCCCGATGACGCCAACTTACAAG 821
Db 110917 GCCTGGATCGCTGCGCGTGTCTGCCACCTTCCCGATGACGCCAACTTACAAG 110858
QY 822 ATGAGACCTCGCTGCGCGTGTCTGCCACCTTCCCGATGACGCCAACTTACAAT 853
Db 110857 ATGAGACCTCGCTGCGCGTGTCTGCCACCTTCCCGATGACGCCAACTTACAAT 110826

```

```

RESULT 8
AC127417 185806 bp DNA linear HTG 15-JUL-2002
LOCUS Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
DEFINITION SEQUENCE, 36 unordered pieces.
AC127417.1 GI:21759524
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 185806)
McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185806)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submission@wustl.edu
----- Project Information -----
Center project name: M.BA0459M02
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173518 bases at least Q40
Consensus quality: 177276 bases at least Q30
Consensus quality: 179611 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 184535; sum-of-contigs
Quality coverage: 3.66 in Q20 bases; agarose-fp
Quality coverage: 3.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 1042: contig of 1042 bp in length
* 1043 1142: gap of unknown length
* 1143 2216: contig of 1074 bp in length
* 2217 2316: gap of unknown length
* 2317 3838: contig of 1522 bp in length
* 3839 3938: gap of unknown length
* 3939 5396: contig of 1458 bp in length
* 5397 5496: gap of unknown length
* 5497 6874: contig of 1378 bp in length
* 6875 6974: gap of unknown length
* 6975 8481: contig of 1507 bp in length
* 8482 8581: gap of unknown length
* 8582 9874: contig of 1293 bp in length
* 9875 9974: gap of unknown length
* 9975 11483: contig of 1509 bp in length
* 11484 11583: gap of unknown length
* 11584 13834: contig of 2251 bp in length
* 13835 13934: gap of unknown length
* 13935 16035: contig of 2101 bp in length
* 16036 16135: gap of unknown length
* 16136 19101: contig of 2965 bp in length
* 19101 19201: gap of unknown length
* 19201 2184: contig of 2984 bp in length
* 2185 22284: gap of unknown length
* 22285 25287: contig of 3003 bp in length
* 25288 25388: gap of unknown length
* 25389 28431: contig of 3044 bp in length
* 28432 28531: gap of unknown length
* 28532 31409: contig of 2878 bp in length
* 31410 31509: gap of unknown length
* 31510 3320: contig of 3811 bp in length
* 3321 35420: gap of unknown length
* 35421 38808: contig of 3388 bp in length
* 38809 38908: gap of unknown length
* 38909 43846: contig of 4938 bp in length
* 43847 43946: gap of unknown length
* 43947 49171: contig of 5225 bp in length
* 49172 49271: gap of unknown length
* 49272 53157: contig of 3886 bp in length
* 53158 53257: gap of unknown length
* 53258 57731: contig of 4474 bp in length
* 57732 57831: gap of unknown length
* 57832 62661: contig of 4829 bp in length
* 62661 62760: gap of unknown length
* 62761 68243: contig of 5483 bp in length
* 68244 68343: gap of unknown length
* 68344 73047: contig of 4704 bp in length
* 73048 73147: gap of unknown length
* 73148 78565: contig of 5418 bp in length
* 78566 78666: gap of unknown length
* 78667 85726: contig of 7061 bp in length
* 85727 85826: gap of unknown length
* 85827 91985: contig of 6159 bp in length
* 91986 92085: gap of unknown length
* 92086 97984: contig of 5899 bp in length
* 97985 98084: gap of unknown length
* 98085 105338: contig of 7254 bp in length
* 105339 105438: gap of unknown length
* 105439 113368: contig of 7930 bp in length
* 113369 113468: gap of unknown length
* 113469 120753: contig of 7285 bp in length
* 120754 120854: gap of unknown length
* 120855 129342: contig of 8489 bp in length
* 129343 129442: gap of unknown length
* 129443 138998: contig of 9556 bp in length
* 138999 139098: gap of unknown length
* 139099 149579: contig of 10481 bp in length
* 149580 149679: gap of unknown length
```

## FEATURES

```
source
1. .185806
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-459M2"
1. 1042
/note="assembly_name:Contig24"
1143. .2216
/note="assembly_name:Contig25"
2317. .3838
/note="assembly_name:Contig26"
3939. .5396
/note="assembly_name:Contig27"
5497. .6874
/note="assembly_name:Contig28"
6975. .8481
/note="assembly_name:Contig29"
8582. .9874
/note="assembly_name:Contig30"
9975. .11483
/note="assembly_name:Contig31"
11584. .13834
/note="assembly_name:Contig32"
13935. .16035
/note="assembly_name:Contig33"
16136. .19100
/note="assembly_name:Contig34"
19201. .22184
/note="assembly_name:Contig35"
22285. .25287
/note="assembly_name:Contig36"
25388. .28431
/note="assembly_name:Contig37"
28532. .31409
/note="assembly_name:Contig38"
31510. .35320
/note="assembly_name:Contig39"
35421. .38808
/note="assembly_name:Contig40"
38909. .43846
/note="assembly_name:Contig41"
43947. .49171
/note="assembly_name:Contig42"
49272. .53157
/note="assembly_name:Contig43"
53258. .57731
/note="assembly_name:Contig44"
57832. .62660
/note="assembly_name:Contig45"
62761. .68243
/note="assembly_name:Contig46"
68344. .73047
/note="assembly_name:Contig47"
73148. .78565
/note="assembly_name:Contig48"
78666. .85726
/note="assembly_name:Contig49"
85827. .91985
/note="assembly_name:Contig50"
92086. .97984
/note="assembly_name:Contig51
clone_end:SP6
vector_side:left"
98085. .105338
/note="assembly_name:Contig52"
105439. .113368
/note="assembly_name:Contig53"
113469. .120753
/note="assembly_name:Contig54"
```

misc\_feature 120854..129342  
 /note="assembly\_name:Contig55"  
 misc\_feature 129413..138998  
 /note="assembly\_name:Contig56"  
 misc\_feature 139099..149579  
 /note="assembly\_name:Contig57"  
 misc\_feature 149680..162343  
 /note="assembly\_name:Contig58"  
 misc\_feature 162444..185806

Query Match 6.3%; Score 92; DB 2; Length 185806;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-42;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCCAACTTACAAG 821  
 |||||  
 DB 159127 GCCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCCAACTTACAAG 159186  
 |||||  
 QY 822 ATCGAGACCTCGCTTGGCCCAACTACAT 853  
 |||||  
 DB 159187 ATCGAGACCTCGCTTGGCCCAACTACAT 159218  
 |||||

RESULT 9  
 AF303002  
 LOCUS AF303002 6123 bp DNA linear PRI 13-NOV-2001  
 DEFINITION Homo sapiens neurogenin 2 gene, partial cds.  
 ACCESSION AF303002  
 VERSION AF303002.1 GI:11875763  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 6123)  
 AUTHORS Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.  
 TITLE Neurogenin2 expression in ventral and dorsal spinal neural tube  
 JOURNAL progenitor cells is regulated by distinct enhancers  
 MEDLINE Dev Biol. 229 (2), 327-339 (2001)  
 PUBMED 21077024  
 REFERENCE 2 (bases 1 to 6123)  
 AUTHORS Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-SEP-2000) Center for Basic Neuroscience - N44.146, UT  
 Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX  
 75390-9111, USA

FEATURES  
 source  
 1..6123  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 mRNA <1..>754  
 CDS <1..754  
 /note="NGN2"  
 /codon\_start=2  
 /product="neurogenin 2"  
 /protein\_id="AA040770.1"  
 /db\_xref="GI:11875764"  
 /translation="ASPALAALTPSSADEEEEEEPGASGGARRORGAEGQGARGG  
 VAAGGCGPRLGLVHDKRRPSRAVSAGKATVQRIKTRRLKANNRERNR  
 MNLNALDALREVLPFPEDAKLKIETLFAHNYIWALETTLRLADHCGGGGLP  
 GALTSEAVLUSFGGSASGDSFSPASTVSCINSPAPSSVSNSTSPISCTLSLSP  
 ASPAGSDMDYWPFPDPDKRYAPHLPIDRCI"  
 BASE COUNT 1484 a 1536 c 1507 g 1596 t  
 ORIGIN  
 |||||

Query Match 2.4%; Score 35; DB 9; Length 6123;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTGGCTTCGCCCAACTACAT 853  
 |||||

DB 383 AAGATCGAGACCTGGCTTCGCCCAACTACAT 417

RESULT 10  
 AC079846\_3/c  
 WPCOMMENT  
 Sequence split into 4 fragments LOCUS AC079846 Accession AC079846  
 Fragment Name Begin End  
 AC079846\_0 1 110000  
 AC079846\_1 100001 210000  
 AC079846\_2 200001 310000  
 AC079846\_3 300001 391531  
 Continuation (4 of 4) of AC079846 from base 300001 (AC079846 Homo sapiens chromosome

Query Match 2.4%; Score 35; DB 2; Length 91531;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTGGCTTCGCCCAACTACAT 853  
 |||||  
 DB 86995 AAGATCGAGACCTGGCTTCGCCCAACTACAT 86961  
 |||||

RESULT 11  
 AC023886/c  
 LOCUS AC023886 179697 bp DNA linear PRI 20-MAR-2002  
 DEFINITION Homo sapiens BAC clone RP11-402J6 from 4, complete sequence.  
 ACCESSION AC023886  
 VERSION AC023886.7 GI:19482381  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 179697)  
 AUTHORS Sulston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 179697)  
 AUTHORS Radionenko,M. and Abbott,A.  
 TITLE The sequence of Homo sapiens BAC clone RP11-402J6  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 179697)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-FEB-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

REFERENCE 4 (bases 1 to 179697)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-MAR-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 5 (bases 1 to 179697)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-MAR-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Mar 15, 2002 this sequence version replaced gi:1735241.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0402J06  
 -----

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC004049, 2000 bp overlap; the clone sequenced to the right is RP11-148B6, Actual end of this clone is at base position 179697 of RP11-402J6.

#### FEATURES

source	
repeat_region	12695..13940 /rpt_family="L1"
repeat_region	13965..14272 /rpt_family="Alu"
repeat_region	14274..14795 /rpt_family="ERV1"
repeat_region	14806..15392 /rpt_family="L1"
repeat_region	21504..22112 /rpt_family="L1"
repeat_region	22131..23166 /rpt_family="L1"
repeat_region	23246..23386 /rpt_family="MaLR"
repeat_region	23511..23654 /rpt_family="MIR"
repeat_region	23655..25486 /rpt_family="L1"
repeat_region	25487..25636 /rpt_family="MIR"
repeat_region	25814..26124 /rpt_family="Alu"
repeat_region	26444..26592 /rpt_family="MIR"
repeat_region	27423..27713 /rpt_family="Alu"
repeat_region	27716..27739 /rpt_family="AT-rich"
repeat_region	27773..27795 /rpt_family="AT-rich"
repeat_region	28020..28429 /rpt_family="ERV1"
repeat_region	29646..29668 /rpt_family="AT-rich"
repeat_region	29727..29871 /rpt_family="L1"
repeat_region	29875..30114 /rpt_family="Alu"
repeat_region	30134..30267 /rpt_family="L1"
repeat_region	30268..30549 /rpt_family="L1"
repeat_region	30550..30883 /rpt_family="MaLR"
repeat_region	32461..32791 /rpt_family="MaLR"
repeat_region	32792..33487 /rpt_family="L1"
repeat_region	33488..33847 /rpt_family="MaLR"
repeat_region	33848..34354 /rpt_family="L1"
repeat_region	34358..34379 /rpt_family="(CATATA)n"
repeat_region	34381..34502 /rpt_family="Alu"
repeat_region	34503..34556 /rpt_family="L1"
repeat_region	34557..34859 /rpt_family="Alu"
repeat_region	34860..35041 /rpt_family="L1"
repeat_region	35219..35676 /rpt_family="L2"
repeat_region	35974..36042 /rpt_family="L2"
repeat_region	36833..37144 /rpt_family="Alu"
repeat_region	36833..36848 /note="match to EST AI537116 (NID:94451251) to15h09.xl"
repeat_region	37147..37410 /rpt_family="Alu"
repeat_region	38154..38424

```

repeat_region      /rpt_family="Alu"
39941..40230
/rpt_family="Alu"
39941..39956
/misc_feature      /note="match to EST AA775240 (NID:g2834574) ad18a05.s1"
repeat_region      40767..40798

Query Match      2.4%; Score 35; DB 9; Length 179697;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGCTTCGCCCAACTACAT 853
|||||
Db 55815 AAGATCGAGACCTCGCTTCGCCCAACTACAT 55781

RESULT 12
HSA133776
LOCUS      HSA133776      1330 bp      DNA      linear      PRI 19-JUN-1999
DEFINITION Homo sapiens gene for neurogenin 3.
ACCESSION AJ133776
VERSION AJ133776.1 GI:5123782
KEYWORDS bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and Mallet,J.
TITLE The human neurogenin 3 homolog maps to chromosome 10q21.3 and its expression pattern is identical to that of its murine counterparts
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1330)
AUTHORS Ravassard,P.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMRC 9923, Hopital de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013 PARIS, FRANCE

FEATURES
source      Location/Qualifiers
1..1330
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene      1..1330
/mrna      /gene="ngn3"
join(<1..157,321..>1330)
/exon      /gene="ngn3"
1..157
/number=1
158..320
/number=2
321..1330
/number=3
322..966
/number=4
/function="early neurogenesis"
/codon_start=1
/product="neurogenin 3"
/protein_id="CA45384.1"
/db_xref="GI:5123783"
/translation="MTQPQSGAPTQVTRERSFPRASEDEVCTSPPTPTPG
NCAAEFGCGRGAPKRLRARGRSRPSKSELALSOKRRRRKANDRRNRHDLNSA
LDALRGVLPFTDDAKLTETLRFAHNYIWAQTQRLADHSLYALEPPAPHCGLG
SPGGPGDWGSLYSPVQAGSLSPAASLEERPGLLGATSSACLSPGSLAFSDFL"
BASE COUNT      230 a 459 c 413 g 228 t
ORIGIN

Query Match      2.2%; Score 32; DB 9; Length 1330;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCACTGAGCAGCAGCGACG 703
|||||
Db 3235 AAGAGCGAGTTGGCACTGAGCAGCAGCGACG 3266

RESULT 13
AF234829
LOCUS      AF234829      5340 bp      DNA      linear      PRI 19-OCT-2001
DEFINITION Homo sapiens neurogenin 3 gene, complete cds.
ACCESSION AF234829
VERSION AF234829.1 GI:13183002
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5340)
AUTHORS del Bosque-Plata,L., Lin,J., Horikawa,Y., Schwarz,P.E., Cox,N.J., Iwasaki,N., Ogata,M., Iwamoto,Y., German,M.S. and Bell,G.I.
TITLE Mutations in the coding region of the neurogenin 3 gene (NEUROG3) are not a common cause of maturity-onset diabetes of the young in Japanese subjects
JOURNAL Japanese subjects
MEDLINE Diabetes 50 (3), 694-696 (2001)
PUBMED 11246894
REFERENCE 2 (bases 1 to 5340)
AUTHORS Lin,J. and German,M.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Hormone Research Institute, University of California San Francisco, 513 Parnassus Ave., San Francisco, CA 94143-0534, USA

FEATURES
source      Location/Qualifiers
1..5340
/organism="Homo sapiens"
/db_xref="taxon:9606"
<3022..>3666
/mrna      /product="neurogenin 3"
3022..3666
/cds      /codon_start=1
/product="neurogenin 3"
/protein_id="AAK15022.1"
/db_xref="GI:13183003"
/translation="MTQPQSGAPTQVTRERSFPRASEDEVCTSPPTPTPG
NCAAEFGCGRGAPKRLRARGRSRPSKSELALSOKRRRRKANDRRNRHDLNSA
LDALRGVLPFTDDAKLTETLRFAHNYIWAQTQRLADHSLYALEPPAPHCGLG
SPGGPGDWGSLYSPVQAGSLSPAASLEERPGLLGATSSACLSPGSLAFSDFL"
BASE COUNT      1215 a 1500 c 1514 g 1111 t
ORIGIN

```

## TITLE

Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

## COMMENT

On Jul 8, 2001 this sequence version replaced g1:14575291.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
<http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
RP11-34333 is from the library RP11-11.2 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/dacpac/home.htm>  
VECTOR: pBACe3.6

## FEATURES

source

1. 165110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-34333"  
/clone\_lib="RP11-11.2"  
7. 147  
/note="THE1C repeat: matches 2. 142 of consensus"  
151. 293  
/note="MIR repeat: matches 2. 148 of consensus"  
1136. 1187  
/note="26 copies 2 mer gt 98% conserved"  
2960. 3272  
/note="AluSg repeat: matches 1. 310 of consensus"  
4144. 4270  
/note="AluSx repeat: matches 1. 134 of consensus"  
4319. 4494  
/note="AluSx repeat: matches 118. 293 of consensus"  
5662. 6438  
/note="L1MEC repeat: matches 272. 1095 of consensus"  
6916. 8080  
/note="L1MEC repeat: matches 1168. 2367 of consensus"  
9934. 10087  
/note="MIR repeat: matches 25. 184 of consensus"  
10353. 10481  
/note="MIR repeat: matches 65. 194 of consensus"  
11077. 11372  
/note="AluSx repeat: matches 1. 294 of consensus"  
11600. 11639  
/note="20 copies 2 mer tc 95% conserved"  
12479. 12608  
/note="65 copies 2 mer at 90% conserved"  
13433. 13876  
/note="MIR repeat: matches 1. 466 of consensus"  
14372. 14576  
/note="AluSg repeat: matches 1. 301 of consensus"  
14577. 14633  
/note="AluSx repeat: matches 243. 299 of consensus"  
14843. 14943  
/note="MER86 repeat: matches 9. 109 of consensus"  
15404. 15613  
/note="MIR repeat: matches 3. 213 of consensus"  
16626. 16705  
/note="MIR repeat: matches 60. 140 of consensus"  
16928. 16993  
/note="33 copies 2 mer gg 66% conserved"  
17476. 17569  
/note="MER81 repeat: matches 2. 114 of consensus"  
17719. 18069  
/note="L1MC4 repeat: matches 7617. 7977 of consensus"  
18117. 18345  
/note="MIR repeat: matches 7. 262 of consensus"  
19169. 19400  
/note="MIR repeat: matches 5. 239 of consensus"  
19436. 19627  
/note="L1MC5 repeat: matches 7728. 7917 of consensus"  
19642. 19716  
/note="MIR repeat: matches 48. 131 of consensus"  
19969. 20010  
/note="21 copies 2 mer tg 100% conserved"  
20623. 20704  
/note="41 copies 2 mer gt 85% conserved"  
20738. 20773  
/note="U2 repeat: matches 1. 36 of consensus"  
20802. 20863  
/note="MIR repeat: matches 1. 62 of consensus"  
20889. 21263  
/note="MIR repeat: matches 188. 541 of consensus"  
21463. 21618  
/note="MIR repeat: matches 46. 192 of consensus"  
22019. 22326  
/note="AluJb repeat: matches 1. 306 of consensus"  
22381. 22564  
/note="L2 repeat: matches 2453. 2629 of consensus"  
22896. 23174  
/note="MIR repeat: matches 117. 413 of consensus"  
23215. 23346  
/note="MIR repeat: matches 48. 188 of consensus"  
23388. 23532  
/note="L2 repeat: matches 2097. 2230 of consensus"  
23533. 23837  
/note="AluJb repeat: matches 1. 303 of consensus"  
23838. 24137  
/note="L2 repeat: matches 1754. 2097 of consensus"  
24291. 24581  
/note="AluSx repeat: matches 1. 300 of consensus"  
24653. 24850  
/note="MIR repeat: matches 1. 200 of consensus"  
26189. 26685  
/note="CpG island"  
/evidence-not\_experimental  
29453. 30918  
/note="CpG island"  
/evidence-not\_experimental  
31084. 31156  
/note="MIR repeat: matches 65. 138 of consensus"  
31618. 31876  
/note="AluJb repeat: matches 29. 275 of consensus"  
32767. 32830  
/note="MIR repeat: matches 76. 139 of consensus"  
33050. 33178  
/note="43 copies 3 mer tcc 72% conserved"  
35112. 36201  
/note="CpG island"  
/evidence-not\_experimental  
36551. 36604  
/note="27 copies 2 mer ac 94% conserved"  
37585. 38254  
/note="L1P15 repeat: matches 5480. 6157 of consensus"  
38265. 38561  
/note="AluSx repeat: matches 1. 300 of consensus"



```

repeat_region 39214..39423
/note="MIR repeat: matches 3. .219 of consensus"
repeat_region 39489..39801
/note="AluX repeat: matches 1. .292 of consensus"
repeat_region 40050..40189
/note="MIR repeat: matches 1. .144 of consensus"
repeat_region 41057..41290
/note="MIR repeat: matches 8. .240 of consensus"
repeat_region 41296..41460
/note="L2 repeat: matches 2569. .2730 of consensus"
repeat_region 41506..41944
/note="L2 repeat: matches 1916. .2416 of consensus"
repeat_region 42388..42698
/note="L2 repeat: matches 1448. .1779 of consensus"
repeat_region 44193..44579
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 44600..44956
/note="LNR16A repeat: matches 90. .445 of consensus"
repeat_region 45240..45300
/note="MER58A repeat: matches 37. .97 of consensus"
repeat_region 45798..45909
/note="LNR41 repeat: matches 90. .192 of consensus"
repeat_region 46826..46871
/note="23 copies 2 mer gt 97% conserved"
repeat_region 50684..50969
/note="AluSq repeat: matches 1. .287 of consensus"
repeat_region 50980..51291
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 52222..52519
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 54065..54260
/note="LIM4 repeat: matches 3865. .4055 of consensus"
repeat_region 54261..54432
/note="FAM repeat: matches 2. .167 of consensus"
repeat_region 54433..54629
/note="LIM4 repeat: matches 3652. .3865 of consensus"
repeat_region 54648..54862
/note="LNR41 repeat: matches 11. .217 of consensus"
repeat_region 54863..55236
/note="MTR1A1 repeat: matches 1. .365 of consensus"
repeat_region 55237..55700
/note="LNR41 repeat: matches 217. .716 of consensus"

Query Match 2.2% Score 32; DB 9; Length 165110;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCAGTGCAGCAGCGAGCG 703
|||||
Db 30112 AAGAGCGAGTTGGCAGTGCAGCAGCGAGCG 30081

RESULT 15
AC021954/c
LOCUS
DEFINITION
Homo sapiens chromosome 10 clone RP11-57E12 map 10; WORKING DRAFT
SEQUENCE, 24 unordered pieces.
ACCESSION
AC021954.3 GI:7417809
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
human.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173341)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Bouckhalter,B., Brown,A., Burkett,G.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5931
Center clone name: 57_E12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 171041; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehotzky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 173341)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F., Boguslavskiy,L., Bouckhalter,B., Brown,A., Burkett,G., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 173341)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F., Boguslavskiy,L., Bouckhalter,B., Brown,A., Burkett,G., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```

* 1 1574: contig of 1574 bp in length
* 1575 1674: gap of 100 bp
* 1675 3043: contig of 1369 bp in length
* 3044 3143: gap of 100 bp
* 3144 5889: contig of 2746 bp in length
* 5890 5989: gap of 100 bp
* 5990 8979: contig of 2990 bp in length
* 8980 9079: gap of 100 bp
* 9080 13674: contig of 4595 bp in length
* 13675 13774: gap of 100 bp
* 13775 18831: contig of 5057 bp in length
* 18832 18931: gap of 100 bp
* 18932 23526: contig of 4595 bp in length
* 23527 23626: gap of 100 bp
* 23627 27386: contig of 3760 bp in length
* 27387 27486: gap of 100 bp
* 27487 32572: contig of 5086 bp in length
* 32573 32672: gap of 100 bp
* 32673 38632: contig of 5960 bp in length
* 38633 38732: gap of 100 bp
* 38733 43735: contig of 5003 bp in length
* 43736 43835: gap of 100 bp
* 43836 49020: contig of 5185 bp in length
* 49021 49120: gap of 100 bp
* 49121 53660: contig of 4540 bp in length
* 53661 53760: gap of 100 bp
* 53761 59544: contig of 5784 bp in length
* 59545 59644: gap of 100 bp
* 59645 66395: contig of 6751 bp in length
* 66396 66495: gap of 100 bp
* 66496 74696: contig of 8201 bp in length
* 74697 74796: gap of 100 bp
* 74797 81326: contig of 6530 bp in length
* 81327 81426: gap of 100 bp
* 81427 89446: contig of 8020 bp in length
* 89447 89546: gap of 100 bp
* 89547 98459: contig of 6913 bp in length
* 98460 98559: gap of 100 bp
* 98560 106498: contig of 9939 bp in length
* 106499 106598: gap of 100 bp
* 106599 116987: contig of 10388 bp in length
* 116987 117086: gap of 100 bp
* 117087 128890: contig of 11804 bp in length
* 128891 128990: gap of 100 bp
* 128991 147290: contig of 18300 bp in length
* 147291 147391: gap of 100 bp
* 147391 173341: contig of 25951 bp in length.

```

## FEATURES

## Source

```

1..173341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/map="10"
/clone="RP11-57E12"
/cloneLib="RPC1-11 Human Male BAC"
1..1574
/note="assembly_fragment"
1675..3043
/note="assembly_fragment"
3144..5889
/note="assembly_fragment"
5990..8979
/note="assembly_fragment"
9080..13674
/note="assembly_fragment"
13775..18831
/note="assembly_fragment"
18932..23526
/note="assembly_fragment"
23627..27386
/note="assembly_fragment"
clone_end:SP6
vector_side:left"

```

```

misc_feature 27487..32572
/note="assembly_fragment"
misc_feature 32673..38632
/note="assembly_fragment"
misc_feature 38733..43735
/note="assembly_fragment"
misc_feature 43836..49020
/note="assembly_fragment"
misc_feature 49121..53660
/note="assembly_fragment"
misc_feature 53761..59544
/note="assembly_fragment"
misc_feature 59645..66395
/note="assembly_fragment"
misc_feature 66496..74696
/note="assembly_fragment"
misc_feature 74797..81326
/note="assembly_fragment"
misc_feature 81427..89446
/note="assembly_fragment"
misc_feature 89547..96459
/note="assembly_fragment"
misc_feature 96560..106498
/note="assembly_fragment"
misc_feature 106599..116986
/note="assembly_fragment"
misc_feature 117087..128890
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 128991..147290
/note="assembly_fragment"
misc_feature 147391..173341
/note="assembly_fragment"
BASE COUNT 44173 a 40928 c 40432 g 45501 t 2307 others
Query Match 2.2% Score 32; DB 2; Length 173341;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 672 AAGACGAGTTGGCACTGAGCAAGCAGCGACG 703
|||||
Db 6552 AAGACGAGTTGGCACTGAGCAAGCAGCGACG 6521
Search completed: April 8, 2003, 22:56:52
Job time : 5752 secs

```

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 18:58:46 ; Search time 374 Seconds

(without alignments)  
8791.223 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460

Sequence: 1 gcaggtagcagagagcag.....agagtacctaaccagtg 1460

Scoring table:

OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 15

Total number of hits satisfying chosen parameters: 4152

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1460	100.0	1491	19	cdNA encoding a no
2	92	6.3	804	19	Mouse neurogenin 3
3	92	6.3	804	21	Murine neurogenin-3
4	92	6.3	861	22	Mouse neurogenin
5	92	6.3	1861	21	Murine neurogenin
6	92	6.3	5567	22	Mouse atonal homol
7	65	4.5	65	24	Rat spliced trans
8	35	2.4	65	24	Mouse spliced tran
9	32	2.2	5340	21	Human neurogenin 3

	10	29	2.0	428	22	AAS33797	Human cdNA encodin
c	11	26	1.8	26	24	ABT03700	Human Neurogenin-3
	12	25	1.7	25	19	AAV42515	PCR primer used to
	13	23	1.6	738	19	AAV27046	Mouse neurogenin 1
	14	23	1.6	738	21	AAZ51977	Murine neurogenin-
	15	23	1.6	790	22	AAF27264	Chicken atonal hom
	16	23	1.6	1332	19	AAV42938	DNA encoding murin
	17	23	1.6	1333	18	AAV74894	Mouse neurogenic d
	18	23	1.6	1385	19	AAV27049	Mouse neurogenin 2
	19	23	1.6	1385	21	AAZ51980	Murine neurogenin-
	20	23	1.6	1385	22	AAF27269	Mouse neurogenin 2
	21	23	1.6	1412	22	AAF27255	Mouse atonal homol
	22	23	1.6	1412	22	AAF27273	Mouse atonal homol
	23	21	1.4	1268	18	AAV74891	Human neurogenic d
	24	21	1.4	1268	19	AAV42932	DNA encoding human
	25	21	1.4	1535	18	AAV74890	Human neurogenic d
	26	21	1.4	1535	19	AAV42931	DNA encoding human
	27	21	1.4	1550	21	AAAG2681	Human NeuroD2 gene
	28	21	1.4	2776	22	AAAL04043	Human reproductive
c	29	21	1.4	2776	22	AAAL04045	Human reproductive
c	30	21	1.4	2776	22	AAK68475	Human immune/haema
	31	21	1.4	2776	22	AAK68476	Human immune/haema
	32	20	1.4	592	24	ABQ49522	Oligonucleotide fo
	33	20	1.4	592	24	ABQ49523	Oligonucleotide fo
	34	19	1.3	50	17	AAT13245	Acetylcholine tran
c	35	19	1.3	50	17	AAT28933	nAChR beta2 subuni
c	36	19	1.3	510	24	ABQ91538	M. capsulatus gene
	37	19	1.3	615	24	ABQ91539	M. capsulatus gene
	38	19	1.3	1260	24	ABQ91700	M. capsulatus gene
c	39	19	1.3	1263	21	AAC43008	Arabidopsis thalia
c	40	19	1.3	1698	21	AAC36256	Arabidopsis thalia
c	41	19	1.3	3424	22	AAAL06199	Human reproductive
c	42	19	1.3	3424	23	ABL98764	Human testicular a
c	43	19	1.3	3425	22	AAAL06200	Human reproductive
c	44	19	1.3	3425	23	ABL98765	Human testicular a
c	45	19	1.3	18733	22	AAK90692	Human immune/haema

#### ALIGNMENTS

##### RESULT 1

AAV42512  
ID AAV42512 standard; cdNA; 1491 BP.

XX AC AAV42512;

XX DT 05-OCT-1998 (first entry)

XX DE cdNA encoding a novel BHLH protein designated RELAX.

XX KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;  
XX KW control; gene expression; transcriptional activator; targeting;  
XX KW protein expression; central nervous system; CNS; treatment;  
XX KW nervous system disorder; ss.

XX OS Rattus sp.

XX FH Key Location/Qualifiers

XX CDS 459..1103

XX FT /\*tag= a

XX FT /product= RELAX

XX PN WO9827206-A2.

XX PD 25-JUN-1998.

XX PF 19-DEC-1997; 97WO-FR02368.

XX PR 19-DEC-1996; 96FR-0015651.

XX PA (RHON ) RHONE-POULENC RORER SA.

PI Mallet J, Ravassard P, Icard-Liepkalns C;

XX WPI: 1998-362775/31.

DR P-PSDB; AAW62991.

XX Basic helix-loop-helix polypeptide and related nucleic acid - with  
PT transcriptional activity, for targeting expression of genes to  
PT central nervous system and treatment of nervous disease

XX Claim 6; Page 20; 28pp; French.

XX The present sequence encodes a basic helix-loop-helix (BHLH) type  
CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.  
CC The protein is used to control and participate in gene expression,  
CC by acting as transcriptional activator, strictly dependent on the  
CC presence of an intact E box (CANNG), particularly for targeting  
CC expression of proteins to the central nervous system (CNS). The  
CC nucleic acid sequence can be used to treat nervous system disorders,  
CC and antisense sequences can be used to control mRNA transcription.

XX Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;

Query Match 100.0%; Score 1460; DB 19; Length 1491;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGTAGCGAGGAGCAGTCCCTGGCCCGCGTGGCTGATTGGCCGCTGGGCACAGGCA 60

DB 1 GCAGGTAGCGAGGAGCAGTCCCTGGCCCGCGTGGCTGATTGGCCGCTGGGCACAGGCA 60

QY 61 GCAGCCCGGAGCAGCCTCTGCTGGGCGGAGGAGCAGATAAAGCGTGGCAGGGGACACA 120

DB 61 GCAGCCCGGAGCAGCCTCTGCTGGGCGGAGGAGCAGATAAAGCGTGGCAGGGGACACA 120

QY 121 CGATTAGCAGCTCAGAAATCCCTCTGGGTCTCAGCACTGCACAGAGCGCGAGACCCCT 180

DB 121 CGATTAGCAGCTCAGAAATCCCTCTGGGTCTCAGCACTGCACAGAGCGCGAGACCCCT 180

QY 181 CCAGCTCTCTTCTGCTCCAGACGCAATTTACTCCAGCGAGGGCGCTCAGAGCTCAG 240

DB 181 CCAGCTCTCTTCTGCTCCAGACGCAATTTACTCCAGCGAGGGCGCTCAGAGCTCAG 240

QY 241 CAAACTTCCAGAGCAGAGAGGGTTCAGTATCCACCGCTCTTACTCTGACCTCTGACCA 300

DB 241 CAAACTTCCAGAGCAGAGAGGGTTCAGTATCCACCGCTCTTACTCTGACCTCTGACCA 300

QY 301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGTAACTTAGAAGCTTCAAGGG 360

DB 301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGTAACTTAGAAGCTTCAAGGG 360

QY 361 TAGAAGAGGGAGTGGTGGCGCTACTCTAGTCCCGGTGAGTGACCTTAAGTCAGAG 420

DB 361 TAGAAGAGGGAGTGGTGGCGCTACTCTAGTCCCGGTGAGTGACCTTAAGTCAGAG 420

QY 421 ACTGTACACCCCTTCCATTTTCCCAACTCAGGAGTGGCGCTCATCTTGGATG 480

DB 421 ACTGTACACCCCTTCCATTTTCCCAACTCAGGAGTGGCGCTCATCTTGGATG 480

QY 481 CGCCCACTTCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGCTCCGACCC 540

DB 481 CGCCCACTTCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGCTCCGACCC 540

QY 541 AAGTGTCTAGTTCCTCAATTTCCACCCCACTAGCCCACTCTCGTACCGAGGAGTGTCCG 600

DB 541 AAGTGTCTAGTTCCTCAATTTCCACCCCACTAGCCCACTCTCGTACCGAGGAGTGTCCG 600

QY 601 AAGCAGAAGCAGTGTACTGCGGAGGACATCGAGGAGCTCCGTGCGCGCGGAGGGC 660

DB 601 AAGCAGAAGCAGTGTACTGCGGAGGACATCGAGGAGCTCCGTGCGCGCGGAGGGC 660

QY 661 GCAACAGGCCAAGAGCGAGTGGCTAGTGAAGCAGCGACGACCGCGCGCAAGAAG 720

DB 661 GCAACAGGCCAAGAGCGAGTGGCTAGTGAAGCAGCGACGACCGCGCGCAAGAAG 720

QY 721 CCAACGACCGGAGGCGCAACCGCATGCAACAACCTTAATCCCGCTGGATGGCTCGCG 780

DB 721 CCAACGACCGGAGGCGCAACCGCATGCAACAACCTTAATCCCGCTGGATGGCTCGCG 780

QY 781 GTGCTCTGCCACCTTCCCGGATGAGCGCAAACTTACAAGATCGAGACCTTCGCTTCG 840

DB 781 GTGCTCTGCCACCTTCCCGGATGAGCGCAAACTTACAAGATCGAGACCTTCGCTTCG 840

QY 841 CCCAACAATACATTTGGGACATGACTCAGACGCTGGCATAGCGGACCAACAGCTTTACG 900

DB 841 CCCAACAATACATTTGGGACATGACTCAGACGCTGGCATAGCGGACCAACAGCTTTACG 900

QY 901 GCCCGAGCGCCCTGTGGCCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGG 960

DB 901 GCCCGAGCGCCCTGTGGCCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGG 960

QY 961 ACTGGGGCTTATCTACTCCCGAGTTTCCCAAGCTGGTAGCTGAGCCGACAGCCTCAT 1020

DB 961 ACTGGGGCTTATCTACTCCCGAGTTTCCCAAGCTGGTAGCTGAGCCGACAGCCTCAT 1020

QY 1021 TGGAGGAGTTCCTGGCTGCAGTGCAGGTCGCCAGCTCCCATCTCTGCTCCCGGGCACCC 1080

DB 1021 TGGAGGAGTTCCTGGCTGCAGTGCAGGTCGCCAGCTCCCATCTCTGCTCCCGGGCACCC 1080

QY 1081 TGGTGTCTCAGACTTCTTGTGAAGGGCCCAACAGGCCCTGGCGGTGGCGCTGGCAG 1140

DB 1081 TGGTGTCTCAGACTTCTTGTGAAGGGCCCAACAGGCCCTGGCGGTGGCGCTGGCAG 1140

QY 1141 AAGGAGGAGTTCAGAGTGTCTGAAATGGAAGGTAGTGGAGGCACTCAGCATCTCGC 1200

DB 1141 AAGGAGGAGTTCAGAGTGTCTGAAATGGAAGGTAGTGGAGGCACTCAGCATCTCGC 1200

QY 1201 CCCTTCTGCTTTCATTTAGTCAGGTCCTGATTTAACCCAGGATTCGACAGTTCTCTTGC 1260

DB 1201 CCCTTCTGCTTTCATTTAGTCAGGTCCTGATTTAACCCAGGATTCGACAGTTCTCTTGC 1260

QY 1261 GCTGTGCGTGCACAAAGGACATTGCGAGGTGATCTCTCTTAACCTCTCTCAGTGTGCC 1320

DB 1261 GCTGTGCGTGCACAAAGGACATTGCGAGGTGATCTCTCTTAACCTCTCTCAGTGTGCC 1320

QY 1321 ACCTCAAACTCCCGCTCCAAAGCAGAGAGCGGTAGCAGTAAATAGTTGGGAGACTCC 1380

DB 1321 ACCTCAAACTCCCGCTCCAAAGCAGAGAGCGGTAGCAGTAAATAGTTGGGAGACTCC 1380

QY 1381 ATACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

DB 1381 ATACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

QY 1441 AGAGTGACCTTAATCCAGTGT 1460

DB 1441 AGAGTGACCTTAATCCAGTGT 1460

RESULT 2

AAV27050

ID AAV27050 standard; cDNA; 804 BP.

XX AAV27050;

AC AAV27050;

XX 17-SEP-1998 (first entry)

DT Mouse neurogenin 3 gene.

XX ds; Mouse; neurogenin; expression vector; recombinant protein;

DE antibody; neurogenesis.

XX Mus sp.

XX Key

FT Location/Qualifiers

FT 160..804

FT /tag= a

FT /product= "Mouse neurogenin 3"

XX WO9813491-A2.  
 PN  
 XX  
 XX PD 02-APR-1998.  
 PD  
 XX PF 24-SEP-1997; 97WO-US17048.  
 XX  
 XX PR 17-SEP-1997; 97US-0932411.  
 PR 27-SEP-1996; 96US-0722570.  
 PR 12-NOV-1996; 96US-0030864.  
 PR 19-DEC-1996; 96US-0772009.  
 XX  
 XX PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 PA  
 XX PI Anderson DJ, Ma Q, Sommer L;  
 PI  
 XX WPI; 1998-230702/20.  
 DR P-PSDB; AAW54947.  
 DR  
 XX PT Mouse neurogenins, useful in neurogenesis - and recombinant nucleic  
 PT acids and proteins derived from rat and xenopus  
 XX  
 XX PS Disclosure; Fig 9; 106pp; English.  
 XX  
 XX CC The Mouse neurogenin 3 is one of several neurogenin proteins discussed  
 CC in the present invention. The neurogenin nucleic acids can be expressed  
 CC in a host cell, transformed using an expression vector, to produce  
 CC recombinant proteins. The proteins and the antibodies raised against  
 CC the proteins are useful in the study of neurogenesis.  
 XX  
 XX CC Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;  
 SQ

Query Match 6.3%; Score 92; DB 19; Length 804;  
 Best Local Similarity 100.0%; Pred. No. 9.3e-35;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGTGTCTGCGGATGCGCCAACTTACAAG 821  
 DB 463 GCGCTGGATGCGCTGCGGTGTCTGCGGATGCGCCAACTTACAAG 522  
 DB 822 ATCGAGACCTGCGCTGCGGATGCGCCAACTTACAT 853  
 DB 523 ATCGAGACCTGCGCTGCGGATGCGCCAACTTACAT 554

RESULT 3  
 AA251981  
 ID AA251981 standard; DNA; 804 BP.  
 XX  
 XX AC AA251981;  
 XX  
 XX DT 04-JUL-2000 (first entry)  
 XX  
 XX DE Murine neurogenin-3 (NGN3) nucleic acid sequence.  
 XX  
 XX KW Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;  
 KW Phox2a protein; neuronal subtype-specific marker; growth factor;  
 KW neural differentiation; transplantation; neuronal dysfunction;  
 KW optical nerve damage; auditory nerve damage, neurodegenerative disorder;  
 KW neuroprotective; nontropic; anticonvulsant; antiparkinsonian; vulnary;  
 KW cerebroprotective; immunosuppressant; antiinfectious; ds.  
 XX  
 XX OS Mus sp.  
 XX  
 XX Key Location/Qualifiers  
 FH 160..804  
 FT CDS /\*tag= a  
 FT /product= "Murine neurogenin-3 protein"  
 FT  
 XX WO200009676-A2.  
 PN  
 XX PD 24-FEB-2000.  
 PD  
 XX

13-AUG-1999; 99WO-US18525.  
 14-AUG-1998; 98US-0096630.  
 (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 Anderson DJ, Lo L;  
 WPI; 2000-256250/22.  
 P-PSDB; AAY70570.  
 Inducing non-neuronal cells to differentiate into neurons and for  
 non-neuronal cells to express a neuronal subtype-specific marker,  
 comprising contacting the non-neuronal cells with a vector containing  
 neurogenin nucleic acid -  
 Claim 1; Fig 1J; 76pp; English.  
 The patent discloses a method for inducing non-neuronal cells (NNC) to  
 differentiate into neurons and for NNCs to express a neuronal subtype  
 -specific marker. Transformed host cells are used as sources of neuronal  
 and other growth factors; in culture for screening compounds that  
 modulate neural differentiation or as sources of recombinantly produced  
 neurogenins and Phox2a proteins for use in transplantation. The cells  
 also have a variety of in vivo uses, e.g. for transplantation at sites of  
 neuronal dysfunction e.g. patients with hearing or vision loss due to  
 optical or auditory nerve damage, brain or spinal cord injuries, and  
 neurodegenerative disorders e.g. Alzheimer's disease. The present  
 sequence encodes murine neurogenin-3 (NGN-3), a transcription factor of  
 NNCs that differentiate into neurons through the recombinant expression of a  
 transcription factor that induces a core program of neurogenesis. Forced  
 expression of murine NGN3 can elicit expression of at least some neuronal  
 phenotypic markers even in NNCs.  
 Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;  
 Query Match 6.3%; Score 92; DB 21; Length 804;  
 Best Local Similarity 100.0%; Pred. No. 9.3e-35;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGTGTCTGCGGATGCGCCAACTTACAAG 821  
 DB 463 GCGCTGGATGCGCTGCGGTGTCTGCGGATGCGCCAACTTACAAG 522  
 QY 822 ATCGAGACCTGCGCTGCGGATGCGCCAACTTACAT 853  
 DB 523 ATCGAGACCTGCGCTGCGGATGCGCCAACTTACAT 554

RESULT 4  
 AAF27266  
 ID AAF27266 standard; cDNA; 861 BP.  
 XX  
 XX AC AAF27266;  
 XX  
 XX DT 24-APR-2001 (first entry)  
 XX  
 XX DE Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.  
 XX  
 XX KW Atonal; homologue; orthologue; atonal-associated protein; deafness;  
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
 KW cellular proliferation; cerebellar granule neuron; gene therapy;  
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;  
 KW transgenic animal; ss.  
 XX  
 XX OS Mus musculus.  
 XX  
 XX PN WO200073764-A2.  
 XX  
 XX PD 07-DEC-2000.  
 XX  
 XX PF 01-JUN-2000; 2000WO-US15410.  
 XX



XX 01-JUN-2000; 2000WO-US15410.  
 PF  
 XX  
 XX 01-JUN-1999; 99US-0137060.  
 PR  
 XX 19-JAN-2000; 2000US-0176993.  
 PR  
 XX  
 XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PA  
 XX  
 XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
 PI  
 XX  
 XX WPI; 2001-032190/04.  
 DR  
 XX P-PSDB; AAB60350.  
 DR  
 XX  
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or  
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
 PT osteoarthritis and abnormal cell proliferation -  
 XX  
 XX  
 XX Disclosure; Page -: 142pp; English.  
 PS  
 XX  
 XX The invention relates to the use of atonal-associated nucleic acid or  
 CC amino acid sequence, or any of its homologues or orthologues as  
 CC therapeutic agents for the treatment of deafness, partial hearing loss,  
 CC vestibular effects due to damage or loss of inner hair cells,  
 CC osteoarthritis and abnormal cell proliferation. The invention also  
 CC encompasses methods of screening for compounds which affect the  
 CC expression of an atonal-associated nucleic acid sequence in an animal,  
 CC and a transgenic animal in which an allele of a native atonal-associated  
 CC gene is replaced by a heterologous nucleic acid sequence, thus  
 CC inactivating the atonal-associated allele. The nucleic acids or proteins  
 CC may be used in a method of treating an animal for hearing impairment,  
 CC joint disease, balance disorders, abnormal cell proliferation, or other  
 CC disease related to loss of a functional atonal-associated nucleic acid or  
 CC protein. They may particularly be used to treat an animal with a  
 CC deficiency in cerebellar granule neurons or their precursors, and may  
 CC also be used in promoting mechanoreceptive cell growth and generating  
 CC hair cells. The present sequence represents an atonal-associated nucleic  
 CC acid sequence referred to in the invention.  
 CC Note: The present sequence is not shown in the specification, but  
 CC was obtained from GenBank.  
 XX  
 XX Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;  
 SQ  
 Query Match 6.3%; Score 92; DB 22; Length 5567;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-35;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 762 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAAG 821  
 Db  
 5226 GCGCTGGATGCGCTGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAAG 5285  
 QY 822 ATCGAGACCTGCGCTTGGCCGACAACTACAT 853  
 Db 5286 ATCGAGACCTGCGCTTGGCCGACAACTACAT 5317  
 RESULT 7  
 ABN31392  
 ID ABN31392 standard; DNA; 65 BP.  
 XX  
 XX AC  
 XX ABN31392;  
 XX  
 XX  
 XX 15-JUL-2002 (first entry)  
 DT  
 XX Rat spliced transcript detection oligonucleotide SEQ ID NO:4140.  
 DE  
 XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
 KW splice variant; transcriptome; oligonucleotide library; ss.  
 XX  
 XX Rattus norvegicus.  
 OS  
 XX WO200210449-A2.  
 PN  
 XX  
 XX 07-FEB-2002.  
 PD

XX 20-JUL-2001; 2001WO-IB01903.  
 PF  
 XX  
 XX 28-JUL-2000; 2000US-221607P.  
 PR  
 XX 02-MAY-2001; 2001US-287724P.  
 PR  
 XX  
 XX (COMP-) COMPUGEN INC.  
 PA  
 XX  
 XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
 PI  
 XX  
 XX WPI; 2002-257383/30.  
 DR  
 XX  
 XX  
 XX New oligonucleotide libraries comprising oligonucleotides which  
 PT selectively hybridize to mRNAs transcribed from a transcription unit of  
 PT a genome, useful for detecting tissue-, pathology-, and  
 PT developmental-specific genes -  
 XX  
 XX  
 XX Example 1; SEQ ID 4140; 47pp; English.  
 PS  
 XX  
 XX The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the  
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises  
 CC several oligonucleotides, each capable of hybridizing selectively to a  
 CC set of messenger RNAs transcribed from a given transcription unit of  
 CC the genome, which encodes one or more messenger RNA splice variants.  
 CC The oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterizing the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcriptomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a  
 CC particular biological or pathological state, and so allowing the  
 CC detection of tissue- and pathology-specific genes such as those genes  
 CC only expressed in specific tissue under a specific pathological  
 CC condition; to detect developmental specific genes; and to detect RNA  
 CC transcripts and splice variants of a transcriptome of a patient suffering  
 CC from a particular disorder. ABN27253 to ABN59589 represent  
 CC oligonucleotide sequences from rats, humans and mice, which are used in  
 CC the exemplification of the present invention.  
 CC .N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 65 BP; 11 A; 20 C; 14 G; 20 T; 0 other;  
 SQ  
 Query Match 4.5%; Score 65; DB 24; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-21;  
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1252 TTCCTTGCTGCTGCGTGCCACAAAGGACATTGCAGGCTGATCTCTCTTACCCCTCCTC 1311  
 Db 1 TTCCTTGCTGCTGCGTGCCACAAAGGACATTGCAGGCTGATCTCTCTTACCCCTCCTC 60  
 QY 1312 AGTGT 1316  
 Db 61 AGTGT 65  
 RESULT 8  
 ABN57521  
 ID ABN57521 standard; DNA; 65 BP.  
 XX  
 XX AC  
 XX ABN57521;  
 XX  
 XX 15-JUL-2002 (first entry)  
 DT  
 XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:30269.  
 DE  
 XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
 KW splice variant; transcriptome; oligonucleotide library; ss.  
 XX  
 XX Mus musculus.  
 OS

```
XX PN WO200210449-A2.
XX PD 07-FEB-2002.
XX PF 20-JUL-2001; 2001WO-IB01903.
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX PA (COMP-) COMPUGEN INC.
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX PI WPI; 2002-257383/30.
XX DR New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes
XX PS Example 1; SEQ ID 30269; 47pp; English.
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridising selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN7253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 65 BP; 16 A; 25 C; 11 G; 13 T; 0 other;

Query Match 2.4%; Score 35; DB 24; Length 65;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 547 TCAGTTCGAATTCACCCACCCAGCCACTGCTC 581
Db 1 TCAGTTCGAATTCACCCACCCAGCCACTGCTC 35

RESULT 9
AAC61089
ID AAC61089 standard; DNA: 5340 BP.
XX AC AAC61089;
XX DT 05-FEB-2001 (first entry)
XX DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
XX KW islet cell precursor identification; diabetes mellitus; human; ds.
XX OS Homo sapiens.
```

```
XX Key Location/Qualifiers
FH CDS 3022..3666
FT /*tag- a "Ngn3"
FT /product- "Neurogenin 3"
FT /note- "Neurogenin 3"

XX WO2000059936-A1.
XX 12-OCT-2000.
XX 28-MAR-2000; 2000WO-US08436.
XX 06-APR-1999; 99US-0128180.
XX (REGC ) UNIV CALIFORNIA.
XX German MS, Lin J;
XX WPI; 2000-664989/64.
XX P-PSDB; AAY85617.
XX Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX PT mellitus and to identify individuals at risk of diabetes -
XX PT Claim 6; Page 46-48; 54pp; English.
XX CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX CC protein AAY85617. The Ngn3 gene is located at chromosome position
XX CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX CC protein sequences, and includes an antibody recognising the Ngn3 protein.
XX CC Also included in the invention is a method for identifying an islet cell
XX CC precursor, the method involves analysing a cell for the expression of the
XX CC Ngn3 gene product, where detection of the product is indicative of an
XX CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX CC reagent for detecting (in a subject) a predisposition to a defect in
XX CC pancreatic islet cell function or formation associated with a defect in
XX CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX CC precursor cells expressing Ngn3, and to alter cellular differentiation in
XX CC culture in vivo to produce new beta-cells to treat patients with diabetes
XX CC mellitus.
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Query Match 2.2%; Score 32; DB 21; Length 5340;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCAGTCGACGACGACG 703
Db 3235 AAGAGCGAGTTGGCAGTCGACGACGACG 3266

RESULT 10
AAS33797
ID AAS33797 standard; CDNA; 428 BP.
XX AC AAS33797;
XX DT 17-DEC-2001 (first entry)
XX DE Human cDNA encoding a novel foetal antigen, SEQ ID No 321.
XX KW Human; foetal tissue antigen; ss; antinflammatory; neuroprotective;
XX KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
XX KW cardiovascular; autoimmune disease; rheumatoid arthritis;
XX KW hyperproliferative disorder; breast neoplasm; cancer;
XX KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX KW cerebral ischaemia; angiogenesis; nervous system disorder;
XX KW Alzheimer's disease; infection; ocular disorder; corneal infection;
XX KW wound healing; epithelial cell proliferation; food additive.
```



[illegible]

```
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-488782/53.
XX P-PSDB; AAU20977.
XX New polynucleotides and polypeptides for diagnosing, treating,
XX preventing or prognosing e.g. diseases or disorders of the nervous,
XX musculoskeletal, excretory, gastrointestinal, reproductive, and
XX respiratory systems
XX Claim 1; SEQ ID No 321; 642pp; English.
XX The invention relates to novel nucleic acids encoding novel human foetal
XX antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
XX by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX are also used in diagnosing a pathological condition or susceptibility
XX to a pathological condition. The antibodies to the antigens can also
XX be used in alleviating symptoms associated with the disorders and in
XX diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbent assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angiodenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX and ocular disorders e.g. corneal infection. The polypeptides can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities, fat content, lipid, protein, carbohydrate, vitamins,
XX minerals, cofactors and other nutritional components. Numerous
XX examples of diseases and disorders treated by the nucleic acids and
XX proteins are given in the specification. The present sequence
XX
XX Query Match 2.0%; Score 29; DB 22; Length 428;
XX Best Local Similarity 100.0%; Pred. No. 0.0007;
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 825 GAGACCTGGCGTTCGCCACACTACAT 853
XX |||||
XX Db 1 GAGACCTGGCGTTCGCCACACTACAT 29
XX
XX RESULT 11
XX ABT03700/c
XX ID ABT03700 standard; DNA; 26 BP.
XX AC ABT03700;
XX DT 13-SEP-2002 (first entry)
XX DE Human Neurogenin-3 gene PCR primer SEQ ID NO: 221.
XX KW Human; cancer; neoplastic disease; tumour specific marker; cytostatic;
XX KW transcription factor; PCR; primer; ss.
XX OS Homo sapiens.
XX PN WO200240716-A2.
XX PD 23-MAY-2002.
XX PF 13-NOV-2001; 2001WO-US43461.
XX PR 16-NOV-2000; 2000US-249508P.
XX PA (CEMI-) CEMINES LLC.
XX PI Palm K;
XX
XX WPI; 2002-537346/57.
XX Determining the presence of neoplastic molecular markers, by
XX identifying the presence of markers in host test sample using array of
XX neoplastic molecular marker specific reagents and analyzing the array
XX of the reagents
XX Example 1; Page 17; 41pp; English.
XX The present invention relates to a method for determining the presence of
XX neoplastic molecular markers in a host, involving the use of neoplastic
XX molecular marker specific reagents to detect such markers and analysing
XX the array of reagents, allowing the identification of the neoplastic
XX disease present. This can be used to determine the best treatment for
XX cancer, in particular neural cell, lung and prostate tumours. The
XX present sequence is a PCR primer useful for detecting the coding
XX sequences of markers of the invention.
XX
XX Sequence 26 BP; 3 A; 10 C; 5 G; 8 T; 0 other;
XX
XX Query Match 1.8%; Score 26; DB 24; Length 26;
XX Best Local Similarity 100.0%; Pred. No. 0.024;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 673 AGAGCGAGTTGGCAGTCAGCAGCAG 698
XX |||||
XX Db 26 AGAGCGAGTTGGCAGTCAGCAGCAG 1
XX
XX RESULT 12
XX AAV42515
XX ID AAV42515 standard; DNA; 25 BP.
XX AC AAV42515;
XX DT 05-OCT-1998 (first entry)
XX DE PCR primer used to isolate part of the RELAX protein coding region.
XX KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
XX KW control; gene expression; transcriptional activator; targeting;
XX KW protein expression; central nervous system; CNS; treatment;
XX KW nervous system disorder; CIG235; PCR primer; ss.
XX OS Synthetic.
XX OS Rattus sp.
XX PN WO9827206-A2.
XX PD 25-JUN-1998.
XX PF 19-DEC-1997; 97WO-FR02368.
XX PR 19-DEC-1996; 96FR-0015651.
XX PA (RHON ) RHONE-POULENC RORER SA.
XX PI Mallet J, Ravassard P, Icard-Lilepkalns C;
XX WPI; 1998-362775/31.
XX Basic helix-loop-helix polypeptide and related nucleic acid - with
XX transcriptional activity, for targeting expression of genes to
XX central nervous system and treatment of nervous disease
XX
XX Example 2; Page 12; 28pp; French.
XX PCR primers AAV42515-16 are used to isolate part of the DNA encoding
XX a basic helix-loop-helix (BHLH) type protein, designated RELAX (Rat
XX Embryonic Longitudinal Axis) protein. The PCR product is termed CIG235.
XX The protein is used to control and participate in gene expression,
XX by acting as transcriptional activator, strictly dependent on the
```

CC presence of an intact E box (CANNTG), particularly for targeting  
 CC expression of proteins to the central nervous system (CNS). The  
 CC nucleic acid sequence can be used to treat nervous system disorders,  
 CC and antisense sequences can be used to control mRNA transcription.

XX Sequence 25 BP; 5 A; 9 C; 6 G; 5 T; 0 other;

Query Match 1.7%; Score 25; DB 19; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.073; Mismatches 0; Indels 0; Gaps 0;

QY 750 AACCTTAAGTCCGCTGGATGGC 774

DB 1 AACCTTAAGTCCGCTGGATGGC 25

RESULT 13

AAV27046

ID AAV27046 standard; cDNA; 738 BP.

XX AC AAV27046;

XX AC AAV27046;

DT 17-SEP-1998 (first entry)

XX Mouse neurogenin 1 gene.

DE ss; Mouse; neurogenin; expression vector; recombinant protein;  
 XX antibody; neurogenesis.  
 XX Mus sp.

XX Key

XX Location/Qualifiers

FT 1..735

FT /\*tag= a

FT /product= "Mouse neurogenin 1"

XX WO9813491-A2.

XX 02-APR-1998.

XX 24-SEP-1997; 97WO-US17048.

XX 17-SEP-1997; 97US-0932411.

XX 27-SEP-1996; 96US-0722570.

XX 12-NOV-1996; 96US-0030864.

XX 19-DEC-1996; 96US-0772009.

XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Ma Q, Sommer L;

XX WPI; 1998-230702/20.

XX P-PSDB; AAW54944.

XX Mouse neurogenins, useful in neurogenesis - and recombinant nucleic  
 XX acids and proteins derived from rat and xenopus

XX Claim 5; Fig 4; 106pp; English.

XX The mouse neurogenin 1 is one of several neurogenin proteins discussed in  
 CC the present invention. The neurogenin nucleic acids can be expressed in  
 CC a host cell, transformed using an expression vector, to produce  
 CC recombinant proteins. The proteins and the antibodies raised against  
 CC the proteins are useful in the study of neurogenesis.

XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

Query Match 1.6%; Score 23; DB 19; Length 738;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACCACT 754

|||||

DB 304 GAGCGCAACCGCATGCACCACT 326

RESULT 14

AAZ51977

ID AAZ51977 standard; DNA; 738 BP.

XX AC AAZ51977;

XX AC AAZ51977;

DT 04-JUL-2000 (first entry)

XX Murine neurogenin-1 (NGN1) nucleic acid sequence.

XX Neurogenin-1; NGN-1; non-neuronal cell; NNC; neurogenesis;

XX Phox2a protein; neuronal subtype-specific marker; growth factor;

XX neural differentiation; transplantation; neuronal dysfunction; disorder;

XX optical nerve damage; auditory nerve damage; neurodegenerative disorder;

XX neuroprotective; nontropic; anticonvulsant; antiparkinsonian; vulnerary;

XX cerebroprotective; immunosuppressant; antineoplastic; ss.

XX Mus sp.

XX Key

XX Location/Qualifiers

FT 1..735

FT /\*tag= a

FT /product= "Murine neurogenin-1 protein"

XX WO200009676-A2.

XX 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18525.

XX 14-AUG-1998; 98US-0096630.

XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Lo L;

XX WPI; 2000-256250/22.

XX P-PSDB; AAY70566.

XX Inducing non-neuronal cells to differentiate into neurons and for

XX non-neuronal cells to express a neuronal subtype-specific marker,

XX comprising contacting the non-neuronal cells with a vector containing

XX neurogenin nucleic acid -

XX Claim 1; Fig 1C; 76pp; English.

XX The patent discloses a method for inducing non-neuronal cells (NNC) to

XX differentiate into neurons and for NNCs to express a neuronal subtype

XX -specific marker. Transformed host cells are used as sources of neuronal

XX and other growth factors; in culture for screening compounds that

XX modulate neural differentiation or as sources of recombinantly produced

XX neurogenins and Phox2a proteins for use in transplantation. The cells

XX also have a variety of in vivo uses, e.g. for transplantation at sites of

XX neuronal dysfunction e.g. patients with hearing or vision loss due to

XX optical or auditory nerve damage, brain or spinal cord injuries, and

XX neurodegenerative disorders e.g. Alzheimer's disease. The present

XX sequence encodes murine neurogenin-1 (NGN-1), a transcription factor.

XX NNCs differentiate into neurons through the recombinant expression of a

XX transcription factor that induces a core program of neurogenesis. Forced

XX expression of murine NGN1 can elicit expression of at least some neuronal

XX phenotypic markers even in NNCs. This can be used in autografting.

XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

Query Match 1.6%; Score 23; DB 21; Length 738;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACCACT 754

|||||

Search completed: April 8, 2003, 21:20:26  
Job time : 404 secs

Db 304 GAGCGCAACCGCATGCACAACT 326

RESULT 15  
AAF27264  
ID AAF27264 standard; cDNA; 790 BP.  
XX  
AC AAF27264;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Chicken atonal homologue ngn2/ath4a cDNA, SEQ ID NO:20.  
XX  
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;  
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
KW cellular proliferation; cerebellar granule neuron; gene therapy;  
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;  
KW transgenic animal; ss.  
XX  
OS Gallus gallus.  
XX  
PN WO200073764-A2.  
XX  
PD 07-DEC-2000.  
XX  
PF 01-JUN-2000; 2000WO-US15410.  
XX  
PR 01-JUN-1999; 99US-0137060.  
PR 19-JAN-2000; 2000US-0176993.  
XX  
XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
PA  
XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
PI  
XX WPI; 2001-032190/04.  
DR  
DR P-PSDB; AAB60357.  
XX  
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or  
PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
PT osteoarthritis and abnormal cell proliferation -  
XX  
PS Disclosure; Page -: 142pp; English.  
XX  
CC The invention relates to the use of atonal-associated nucleic acid or  
CC amino acid sequence, or any of its homologues or orthologues as  
CC therapeutic agents for the treatment of deafness, partial hearing loss,  
CC vestibular effects due to damage or loss of inner hair cells,  
CC osteoarthritis and abnormal cell proliferation. The invention also  
CC encompasses methods of screening for compounds which affect the  
CC expression of an atonal-associated nucleic acid sequence in an animal,  
CC and a transgenic animal in which an allele of a native atonal-associated  
CC gene is replaced by a heterologous nucleic acid sequence, thus  
CC inactivating the atonal-associated allele. The nucleic acids or proteins  
CC may be used in a method of treating an animal for hearing impairment,  
CC joint disease, balance disorders, abnormal cell proliferation, or other  
CC disease related to loss of a functional atonal-associated nucleic acid or  
CC protein. They may particularly be used to treat an animal with a  
CC deficiency in cerebellar granule neurons or their precursors, and may  
CC also be used in promoting mechanoreceptive cell growth and generating  
CC hair cells. The present sequence represents an atonal-associated nucleic  
CC acid sequence referred to in the invention.  
CC Note: The present sequence is not shown in the specification, but  
CC was obtained from GenBank.  
XX  
SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 790;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACT 754  
|||||  
Db 374 GAGCGCAACCGCATGCACAACT 396  
|||||

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 21:09:53 ; Search time 164 Seconds  
(without alignments)  
7808.933 Million cell updates/sec

Title: US-09-595-947C-1  
Perfect score: 1460  
Sequence: 1 gcaggtagcagagagagcag.....agagtgcacctaaccagtgt 1460

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 593429 seqs, 438583890 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1316

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	92	6.3	645	9	US-10-004-717-4
2	92	6.3	861	9	US-10-004-717-24
3	92	6.3	1861	10	US-09-817-360-3
4	32	2.2	5340	10	US-09-817-360-1
5	23	1.6	738	7	US-08-722-570-13
6	23	1.6	790	9	US-10-004-717-20
7	23	1.6	1385	9	US-10-004-717-30
8	23	1.6	1412	9	US-10-004-717-6
9	23	1.6	1412	9	US-10-004-717-37
10	20	1.4	352	10	US-09-728-445-108
11	19	1.3	500	10	US-09-783-590-2157
12	19	1.3	1263	9	US-09-938-842A-1036
13	18	1.2	4797	10	US-09-751-797-25
14	18	1.2	20272	10	US-09-908-711-145
15	18	1.2	25619	10	US-09-908-711-143
16	18	1.2	25619	10	US-09-764-898-302
17	18	1.2	31728	9	US-10-114-170-64
18	18	1.2	402850	9	US-09-844-653-5
19	17	1.2	31	10	US-09-817-360-17

20	17	1.2	232	10	US-09-960-352-14228
21	17	1.2	285	10	US-09-960-352-6887
22	17	1.2	285	10	US-09-960-352-8778
23	17	1.2	307	10	US-09-783-590-12149
24	17	1.2	454	9	US-10-040-739-1094
25	17	1.2	454	10	US-09-864-761-257
26	17	1.2	469	10	US-09-864-761-5374
27	17	1.2	471	10	US-09-864-761-11
28	17	1.2	473	10	US-09-864-761-22148
29	17	1.2	474	10	US-09-864-761-353
30	17	1.2	474	10	US-09-864-761-10932
31	17	1.2	487	10	US-09-864-761-3320
32	17	1.2	495	10	US-09-864-761-5673
33	17	1.2	499	10	US-09-783-590-4421
34	17	1.2	594	10	US-09-864-761-7938
35	17	1.2	756	10	US-09-910-943-147
36	17	1.2	1074	9	US-10-004-717-18
37	17	1.2	1238	9	US-09-991-496-75
38	17	1.2	1238	10	US-09-874-923-75
39	17	1.2	1362	9	US-09-970-966-208
40	17	1.2	1362	10	US-09-825-294-208
41	17	1.2	1422	10	US-09-815-242-4334
42	17	1.2	1425	10	US-09-815-242-8558
43	17	1.2	1669	10	US-09-880-107-2159
44	17	1.2	1747	9	US-09-764-868-1334
45	17	1.2	1752	9	US-09-764-868-137

ALIGNMENTS

RESULT 1  
US-10-004-717-4  
; Sequence 4, Application US/10004717  
; Publication No. US20020192665A1  
; GENERAL INFORMATION:  
; APPLICANT: ZOGHBI, HUDA Y.  
; APPLICANT: YANG, QI  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
; FILE REFERENCE: P01899054  
; CURRENT APPLICATION NUMBER: US/10/004,717  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: 09/585,645  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/176,993  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: 60/137,060  
; PRIOR FILING DATE: 1999-06-01  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 645  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-004-717-4

Query Match	6.3%	Score 92;	DB 9;	Length 645;
Best Local Similarity	100.0%;	Pred. No. 5.5e-39;		
Matches	92;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	762	GGCGTGGATCGCTCGGGGTGCTCTGCCACCTTCCCGGATGAGCGCCAACTTACAAG	821	
Db	304	GGCGTGGATCGCTCGGGGTGCTCTGCCACCTTCCCGGATGAGCGCCAACTTACAAG	363	
QY	822	ATCGAGACCTCGCTCGGGGTGCTCTGCCACCTTCCCGGATGAGCGCCAACTTACAAG	853	
Db	364	ATCGAGACCTCGCTCGGGGTGCTCTGCCACCTTCCCGGATGAGCGCCAACTTACAAG	395	
RESULT 2				
US-10-004-717-24				

; Sequence 24, Application US/10004717  
; Publication No. US20020192665A1  
; GENERAL INFORMATION:  
; APPLICANT: ZOGHBI, HUDA Y.  
; APPLICANT: YANG, QI  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
; FILE REFERENCE: P01899054  
; CURRENT APPLICATION NUMBER: US/10/004,717  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: 09/585,645  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/176,993  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: 60/137,060  
; PRIOR FILING DATE: 1999-06-01  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 861  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-004-717-24

Query Match 6.3%; Score 92; DB 9; Length 861;  
Best Local Similarity 100.0%; Pred. No. 5.4e-39;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 762 GCCTGGATCGCTGGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821  
DB 463 GCCTGGATCGCTGGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 522  
QY 822 ATCGAGACCTCGCTTCGCCCACTACAT 853  
DB 523 ATCGAGACCTCGCTTCGCCCACTACAT 554

RESULT 3  
US-09-817-360-3  
; Sequence 3, Application US/09817360  
; Patent No. US20020015696A1  
; GENERAL INFORMATION:  
; APPLICANT: German, Michael S.  
; APPLICANT: Lin, Joseph  
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS  
; TITLE OF INVENTION: AND DELIVERY OF INSULIN  
; FILE REFERENCE: UCSF-129CIP  
; CURRENT APPLICATION NUMBER: US/09/817,360  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: 09/535,145  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: 60/128,180  
; PRIOR FILING DATE: 1999-04-06  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1861  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-817-360-3

Query Match 6.3%; Score 92; DB 10; Length 1861;  
Best Local Similarity 100.0%; Pred. No. 5e-39;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 762 GCCTGGATCGCTGGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821  
DB 1396 GCCTGGATCGCTGGCGGTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 1455  
QY 822 ATCGAGACCTCGCTTCGCCCACTACAT 853  
DB 1456 ATCGAGACCTCGCTTCGCCCACTACAT 1487

RESULT 4  
US-09-817-360-1  
; Sequence 1, Application US/09817360  
; Patent No. US20020015696A1  
; GENERAL INFORMATION:  
; APPLICANT: German, Michael S.  
; APPLICANT: Lin, Joseph  
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS  
; TITLE OF INVENTION: AND DELIVERY OF INSULIN  
; FILE REFERENCE: UCSF-129CIP  
; CURRENT APPLICATION NUMBER: US/09/817,360  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: 09/535,145  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: 60/128,180  
; PRIOR FILING DATE: 1999-04-06  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 5340  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-09-817-360-1

Query Match 2.2%; Score 32; DB 10; Length 5340;  
Best Local Similarity 100.0%; Pred. No. 4.9e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 672 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 703  
DB 3235 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 3266

RESULT 5  
US-08-722-570-13  
; Sequence 13, Application US/08722570  
; Publication No. US20030044887A1  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David J.  
; APPLICANT: Ma, Qifu  
; TITLE OF INVENTION: NEUROGENIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/722,570  
; FILING DATE: 27-SEP-1996  
; CLASSIFICATION: 5365  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 738 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown

; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
US-08-722-570-13

Query Match 1.6%; Score 23; DB 7; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.038; 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
|||||  
Db 304 GAGCGCAACCGCATGCACACCT 326

## RESULT 6

US-10-004-717-20  
; Sequence 20, Application US/10004717  
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 20

; LENGTH: 790

; TYPE: DNA

; ORGANISM: chicken

US-10-004-717-20

## Query Match

Best Local Similarity 1.6%; Score 23; DB 9; Length 790;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
|||||  
Db 374 GAGCGCAACCGCATGCACACCT 396

## RESULT 7

US-10-004-717-30

; Sequence 30, Application US/10004717  
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 30

; LENGTH: 1385

; TYPE: DNA

; ORGANISM: Mus musculus  
US-10-004-717-30

Query Match 1.6%; Score 23; DB 9; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
|||||  
Db 742 GAGCGCAACCGCATGCACACCT 764

## RESULT 8

US-10-004-717-6

; Sequence 6, Application US/10004717  
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 1412

; TYPE: DNA

; ORGANISM: Mus musculus

US-10-004-717-6

## Query Match

Best Local Similarity 1.6%; Score 23; DB 9; Length 1412;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
|||||  
Db 445 GAGCGCAACCGCATGCACACCT 467

## RESULT 9

US-10-004-717-37

; Sequence 37, Application US/10004717  
; Publication No. US20020192665A1

; GENERAL INFORMATION:

; APPLICANT: ZOGHBI, HUDA Y.

; APPLICANT: YANG, QI

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,

; FILE REFERENCE: P01899US4

; CURRENT APPLICATION NUMBER: US/10/004,717

; PRIOR FILING DATE: 2002-08-16

; PRIOR APPLICATION NUMBER: 09/585,645

; PRIOR FILING DATE: 2000-06-01

; PRIOR APPLICATION NUMBER: 60/176,993

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: 60/137,060

; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 37

; LENGTH: 1412

; TYPE: DNA

; ORGANISM: Mus musculus

US-10-004-717-37

Query Match 1.48; Score 23; DB 9; Length 1412;  
 Best Local Similarity 100.0%; Pred. No. 0.036;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACT 754  
 Db 445 GAGCGCAACCGCATGCACAACT 467

RESULT 10

US-09-728-445-108/c  
 ; Sequence 108, Application US/09728445  
 ; Patent No. US20020102543A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Friedrich, Glenn  
 ; APPLICANT: Zambrowicz, Brian  
 ; APPLICANT: Sands, Arthur T.  
 ; TITLE OF INVENTION: No. US20020102543A1 Mutated Mammalian Cells and  
 ; TITLE OF INVENTION: Animals  
 ; FILE REFERENCE: LEX-0102-USA  
 ; CURRENT APPLICATION NUMBER: US/09/728,445  
 ; CURRENT FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/168,358  
 ; PRIOR FILING DATE: 1999-12-01  
 ; NUMBER OF SEQ ID NOS: 891  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 108  
 ; LENGTH: 352  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 US-09-728-445-108

Query Match 1.48; Score 20; DB 10; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1010 CACAGCCTCATGGAGGAGT 1029  
 Db 209 CACAGCCTCATGGAGGAGT 190

RESULT 11

US-09-783-590-2157/c  
 ; Sequence 2157, Application US/09783590  
 ; Patent No. US2002010850A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dillon, Patrick J.  
 ; APPLICANT: Haseltine, William A.  
 ; APPLICANT: Li, Haodong  
 ; APPLICANT: Rosen, Craig A.  
 ; APPLICANT: Ruben, Steven M.  
 ; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
 ; FILE REFERENCE: PO-416.2C1  
 ; CURRENT APPLICATION NUMBER: US/09/783,590  
 ; CURRENT FILING DATE: 2000-02-15  
 ; PRIOR APPLICATION NUMBER: 08/420,856  
 ; PRIOR FILING DATE: 1995-04-12  
 ; PRIOR APPLICATION NUMBER: 08/346,731  
 ; PRIOR FILING DATE: 1994-11-21  
 ; NUMBER OF SEQ ID NOS: 12485  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2157  
 ; LENGTH: 500  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (84)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (207)

; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (215)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (324)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (354)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (360)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (368)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (369)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (379)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (393)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (411)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (418)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (439)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (440)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (441)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; NAME/KEY: misc feature  
 ; LOCATION: (451)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 US-09-783-590-2157

Query Match 1.38; Score 19; DB 10; Length 500;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1319 CCACCTCAAACTCCGCTC 1337  
 Db 23 CCACCTCAAACTCCGCTC 5

RESULT 12

US-09-938-842A-1036/c  
 ; Sequence 1036, Application US/09938842A  
 ; Patent No. US20020160378A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Kreps, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINI  
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938,842A  
 ; CURRENT FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111



;; PRIOR FILING DATE: 2001-06-22  
;; NUMBER OF SEQ ID NOS: 5379  
;; SEQ ID NO 1036  
;; LENGTH: 1263  
;; TYPE: DNA  
;; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1036

Query Match 1.3%; Score 19; DB 9; Length 1263;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 GCGCAACAGGCCCAAGAG 676  
|||||  
DB 599 GCGCAACAGGCCCAAGAG 581

RESULT 13  
US-09-751-797-25  
; Sequence 25, Application US/09751797  
; Patent No. US20010024652A1  
; GENERAL INFORMATION:  
; APPLICANT: Dumoutier, Laure  
; APPLICANT: Loubet, Jamila  
; APPLICANT: Renaud, Jean-Christophe  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Factors  
; TITLE OF INVENTION: (Tifs) The Proteins Encoded, and Uses Thereof  
; FILE REFERENCE: LUD 5543.2  
; CURRENT APPLICATION NUMBER: US/09/751,797  
; CURRENT FILING DATE: 2000-12-29  
; PRIOR APPLICATION NUMBER: 09/419,568  
; PRIOR FILING DATE: 1999-10-18  
; PRIOR APPLICATION NUMBER: US09/178,973  
; PRIOR FILING DATE: 1998-10-26  
; NUMBER OF SEQ ID NOS: 29  
; SEQ ID NO 25  
; LENGTH: 4797  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-751-797-25

Query Match 1.2%; Score 18; DB 10; Length 4797;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1152 GTCAGAGCTGCTGAAAT 1169  
|||||  
DB 3460 GTCAGAGCTGCTGAAAT 3477

RESULT 14  
US-09-908-711-145/c  
; Sequence 145, Application US/09908711  
; Patent No. US20020045230A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA128  
; CURRENT APPLICATION NUMBER: US/09/908,711  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US01/01360  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,867  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01344  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01345  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,888  
; PRIOR FILING DATE: 2001-01-17

;; PRIOR APPLICATION NUMBER: US01/01329  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,905  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01354  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,891  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01339  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,869  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01340  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,874  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01334  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,898  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01320  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,853  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01349  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,902  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01239  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,870  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01348  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,882  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01347  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,896  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01307  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,864  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01341  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,856  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01336  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 09/764,868  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: US01/01312  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 60/179,065  
;; PRIOR FILING DATE: 2000-01-31  
;; PRIOR APPLICATION NUMBER: 60/180,628  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: 60/209,467  
;; PRIOR FILING DATE: 2000-06-07  
;; NUMBER OF SEQ ID NOS: 167  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 145  
;; LENGTH: 20272  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-908-711-145

Query Match 1.2%; Score 18; DB 10; Length 20272;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1153 TCAGAGCTGCTGAAATG 1170

Db 6642 TCAGAGCTGCTGAAATG 6625  
|||||

RESULT 15  
US-09-908-711-143/c  
; Sequence 143, Application US/09908711  
; Patent No. US20020045230A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA128  
; CURRENT APPLICATION NUMBER: US/09/908,711  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US01/01360  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,867  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01344  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01345  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,888  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01329  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,905  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01354  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01339  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,869  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01340  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,874  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01334  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,898  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01320  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,853  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01349  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,902  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01239  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,870  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01348  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,882  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01347  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,896  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01307  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,864  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01341  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,856

; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01336  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,868  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01312  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 167  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 143  
; LENGTH: 25619  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-711-143  
Query Match 1.2%; Score 18; DB 10; Length 25619;  
Best Local Similarity 100.0%; Pred.No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1153 TCAGAGCTGCTGAAATG 1170  
|||||  
Db 1048 TCAGAGCTGCTGAAATG 1031  
Search completed: April 8, 2003, 23:20:39  
Job time : 285 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 20:42:04 ; Search time 2260 Seconds  
(without alignments)  
10462.571 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460  
Sequence: 1 gcaggtacgagagagcag.....agagtacctaaccagtgt 1460

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 15

Total number of hits satisfying chosen parameters: 29042

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	6.3	593	17	AZ296526
2	60	4.1	1540	11	AK008017
3	23	1.6	600	12	BG808248
4	23	1.6	804	9	AL540071
5	23	1.6	814	14	BQ178789
6	23	1.6	1269	12	BG854922

7	1.4	319	12	BE936551
8	1.4	464	17	AQ753599
9	1.4	536	17	AZ508899
10	1.4	718	10	BE263765
11	1.4	829	13	BI910102
12	1.4	913	12	BF204175
13	1.4	953	9	AU067624
14	1.4	962	14	BQ686909
15	1.4	1022	13	BM563662
16	1.4	1039	12	BE780690
17	1.4	1042	12	BG419220
18	1.4	241	9	AA113743
19	1.4	351	14	T70947
20	1.4	397	13	BI445873
21	1.4	444	10	AV944801
22	1.4	476	13	BM253695
23	1.4	550	10	AV939670
24	1.4	571	17	CNS01YSD
25	1.4	971	17	CNS02YD5
26	1.4	1075	14	BM810917
27	1.3	201	10	BB071548
28	1.3	234	10	BB269747
29	1.3	260	10	BE068280
30	1.3	301	10	BB305111
31	1.3	366	10	AW426722
32	1.3	370	10	AW481542
33	1.3	468	17	AZ286451
34	1.3	496	12	BF191528
35	1.3	507	12	BF408307
36	1.3	529	17	AZ830930
37	1.3	533	17	AZ985482
38	1.3	541	9	AU057600
39	1.3	552	13	BM030263
40	1.3	555	10	BE600721
41	1.3	564	13	BM088009
42	1.3	570	17	AQ594075
43	1.3	577	10	AW672224
44	1.3	579	10	AV597322
45	1.3	585	17	AQ631359

#### ALIGNMENTS

RESULT 1  
AZ296526/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AZ296526 593 bp DNA linear GSS 27-JUL-2000  
RPCI-23-160G18-TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18  
, DNA sequence.  
AZ296526 GI:9538311  
GSS.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 593)  
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret  
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroi,M., de Jong,P.  
and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from



```

/strain="C57BL/6J"
/db_xref="FANTOM.DB:2010001M19"
/db_xref="MGI:1907403"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..1540
/gene="Atoh5"
241..885
/gene="Atoh5"
/feature="atonal homolog 5 (Drosophila)"
data source:MGI, source key:MGI:893591, evidence:ISS
putative"
/codon_start=1
/protein_id="BAB25411.1"
/db_xref="GI:12841942"
/db_xref="MGI:893591"
/translation="MAPHLDALTIQVSPETQPPGASDHEVLSSNTPPSHTLIPR
DCSEAGDCRGTSRKLRRGRNRPKSELAKRRSRKRRKANDRRNRNHNLSA
LDALRGVLPDDAKLKVETLRFHNYIWAQTQTLRIADHSLGPPVPCGELGS
PGGSGNGDWGSIYSPVSOAGNLSPASLEEFGLQVPSPSYLLPGLVFSDFL"

BASE COUNT      306 a 485 c 397 g 352 t
ORIGIN
Query Match      4.1%; Score 60; DB 11; Length 1540;
Best Local Similarity 100.0%; Pred. No. 4.8e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCAACCTTACAAG 821
|||||
DB 544 GCGCTGGATGCGTGGCGGTGTCCTGCCACCTTCCCGATGACGCCAACCTTACAAG 603

RESULT 3
BG808248
LOCUS      600 bp mRNA linear EST 20-DEC-2001
DEFINITION      2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
mRNA sequence.
ACCESSION      BG808248
VERSION      BG808248.1 GI:17955225
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)
AUTHORS      Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
TITLE      Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
JOURNAL      Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE      21671825
COMMENT      Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

FEATURES
source
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Manniat); Cloning Technique: CUA Cloning (cloneAmp,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGCTCACTGAATCTGAGTG---. Other

BASE COUNT      154 a 270 c 278 g 99 t
ORIGIN
Query Match      1.6%; Score 23; DB 9; Length 804;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CTGCGCTTCGCCACCACTACAT 853
|||||
DB 770 CTGCGCTTCGCCACCACTACAT 792

RESULT 5
BQ178789
LOCUS      814 bp mRNA linear EST 30-APR-2002
DEFINITION      BQ178789
IMAGE:5701758 5', mRNA sequence.
ACCESSION      BQ178789
VERSION      BQ178789.1 GI:20354281

BASE COUNT      94 a 238 c 161 g 107 t
ORIGIN
Query Match      1.6%; Score 23; DB 12; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 4 GAGCGCAACCGCATGCACACCT 26

RESULT 4
AL540071
LOCUS      804 bp mRNA linear EST 16-FEB-2001
DEFINITION      AL540071 LTI_FLO13_FBrn1 Homo sapiens cDNA clone CS0DF035YD19 5
prime, mRNA sequence.
ACCESSION      AL540071
VERSION      AL540071.1 GI:12869886
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 804)
AUTHORS      Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF035YD19"
/clone_lib="LTI_FLO13_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      154 a 270 c 278 g 99 t
ORIGIN
Query Match      1.6%; Score 23; DB 9; Length 804;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CTGCGCTTCGCCACCACTACAT 853
|||||
DB 770 CTGCGCTTCGCCACCACTACAT 792

RESULT 5
BQ178789
LOCUS      814 bp mRNA linear EST 30-APR-2002
DEFINITION      BQ178789
IMAGE:5701758 5', mRNA sequence.
ACCESSION      BQ178789
VERSION      BQ178789.1 GI:20354281

```

Information regarding entire library may be found at  
[http://pga.swmed.edu/Data/Libraries/microarray\\_cdna\\_libraries.htm](http://pga.swmed.edu/Data/Libraries/microarray_cdna_libraries.htm).

# KEYWORDS SOURCE

EST.  
house musculus  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 814)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Dr. James Lin, University of Iowa  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

# FEATURES source

Seq primer: pYX-5.  
Location/Qualifiers  
1..814  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5701758"  
/clone\_lib="NIH\_BMAP\_EV0"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 15.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; the library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
, is GGGCGGGA. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP): "Gene Discovery in the Developing Mouse Nervous  
System", supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."  
151 a 268 c 291 g 102 t 2 others

# BASE COUNT

ORIGIN  
Query Match 1.6%; Score 23; DB 14; Length 814;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACT 754  
|||||||w|||||||  
Db 663 GAGCGCAACCGCATGCACAACT 685

# RESULT 6

BG854922 1269 bp mRNA linear EST 29-MAY-2001  
LOCUS 1024041C06.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II  
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION BG854922  
VERSION BG854922.1 GI:14236106  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadales;  
1 (bases 1 to 1269)  
Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,  
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.

# TITLE

Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 2  
Unpublished (2000)  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: [chauser@duke.edu](mailto:chauser@duke.edu).

# FEATURES source

Location/Qualifiers  
1..1269  
/organism="Chlamydomonas reinhardtii"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap  
II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; This library, constructed by John Davies and Jeffrey  
McDermott, combines cDNAs from CC-1690 cells grown to  
mid-log phase in TAP (acetate-containing) medium in the  
light, TAP medium in the dark, HS (minimal) medium in  
ambient levels of CO2 and HS medium bubbled with 5% CO2.  
polyA mRNA was purified from each sample, pooled and cDNA  
synthesized. The cDNA was directionally cloned into lambda  
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
ZAP clones by superinfection with ExAssist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bonaldo et al (1996) Genome Research 6: 791-806."  
558 a 231 c 444 g 12 t 24 others

# Query Match

Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 653 CGGAGGGCGCAACAGGCCCAAGA 675  
|||||||w|||||||  
Db 1244 CGGAGGGCGCAACAGGCCCAAGA 1266

# RESULT 7

BE936551 319 bp mRNA linear EST 02-OCT-2000  
LOCUS RCS-NF0053-310800-024-E01 NT0053 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE936551  
VERSION BE936551.1 GI:10462627  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 319)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordip,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

# TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2=RC5-NT0053-310>)  
800-024-E01et3-2000-08-31&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 319.

## FEATURES

source  
1. .319  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NT0053"  
/dev\_stage="Adult"  
/note="Organ: nervous tumor; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
53 a 87 c 106 g 73 t

## BASE COUNT

ORIGIN  
Query Match 1.4%; Score 21; DB 12; Length 319;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCGTGGCGGTGGGC 1132

Db 280 AACAGGCCGTGGCGGTGGGC 300

## RESULT 8

AQ753599 464 bp DNA linear GSS 22-JUL-1999  
LOCUS HS\_2117\_A2\_H03\_MR CIT Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone Plate-2117 Col-6 Row-O, DNA sequence.

ACCESSION AQ753599

VERSION AQ753599.1 GI:5578650

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 464)

REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 2117 row: 0 column: 6

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 464.

## FEATURES

source  
1. .464

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Plate-2117 Col-6 Row-O"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 125 a 114 c 101 g 123 t 1 others  
ORIGIN

Query Match 1.4%; Score 21; DB 17; Length 464;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 CAAAGTGTCCCAAGAGACCCAG 512

Db 106 CAAAGTGTCCCAAGAGACCCAG 126

## RESULT 9

AZ506899

LOCUS

DEFINITION

IM0348G13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0348G13 F, DNA sequence.

ACCESSION AZ506899

VERSION AZ506899.1 GI:10688215

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 536)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

CONTACT: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [rdunn@genetics.utah.edu](mailto:rdunn@genetics.utah.edu)

Insert Length: 10000

Plate: 0348 row: G column: 13

Seq primer: CGTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 536.

Location/Qualifiers

1. .536

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0348G13"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of PWD42 (g11473214|gb|AF129072.1), a copy-number

inducible derivative of plasmid RL. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance.

```

BASE COUNT      156 a      119 c      81 g      180 t
ORIGIN

Query Match
Best Local Similarity 1.4%; Score 21; DB 17; Length 536;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 TGCCCTGTGGGAGCTGGGAA 936
|||||
Db 406 TGCCCTGTGGGAGCTGGGAA 426

RESULT 10
BE263765
LOCUS 601194122F1 NIH_MGC_7 Homo sapiens cdna clone IMAGE:5337940 5',
DEFINITION mRNA sequence.
ACCESSION BE263765
VERSION BE263765.1 GI:9137311
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 718)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Plate: LICM222 row: 1 column: 05
High quality sequence stop: 649.
FEATURES
source
location/Qualifiers
1..718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3537940"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      115 a      195 c      260 g      148 t
ORIGIN

Query Match
Best Local Similarity 1.4%; Score 21; DB 10; Length 718;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGC 1132
|||||
Db 536 AACAGGCCCTGGCGGTGGC 556

RESULT 11
BI910102
LOCUS B1910102 829 bp mRNA linear EST 16-OCT-2001
DEFINITION 603067946F1 NIH_MGC_118 Homo sapiens cdna clone IMAGE:5217154 5',
mRNA sequence.
ACCESSION BI910102
VERSION BI910102.1 GI:16173443
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 913)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cdna Library Preparation: Ling Hong/Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM996 row: a column: 07
High quality sequence stop: 714.
FEATURES
source
location/Qualifiers
1..913
/organism="Homo sapiens"

```

```

REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cdna Library Preparation: Life Technologies, Inc.
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM11545 row: k column: 11
High quality sequence stop: 828.
FEATURES
source
location/Qualifiers
1..829
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5217154"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH10B"
/Note="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 027. Note:
this is a NIH_MGC Library."
BASE COUNT      146 a      224 c      293 g      166 t
ORIGIN

Query Match
Best Local Similarity 1.4%; Score 21; DB 13; Length 829;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGC 1132
|||||
Db 550 AACAGGCCCTGGCGGTGGC 570

RESULT 12
BF204175
LOCUS 601867625F1 NIH_MGC_17 Homo sapiens cdna clone IMAGE:4110222 5',
DEFINITION mRNA sequence.
ACCESSION BF204175
VERSION BF204175.1 GI:11097761
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 913)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cdna Library Preparation: Ling Hong/Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM996 row: a column: 07
High quality sequence stop: 714.
FEATURES
source
location/Qualifiers
1..913
/organism="Homo sapiens"

```





KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1022)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LCM1916 row: k column: 15  
High quality sequence stop: 654.  
Location/Qualifiers  
1..1022  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5441798"  
/clone\_lib="NIH\_MGC\_98"  
/tissue\_type="astrocytoma grade IV, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

BASE COUNT 177 a 324 c 319 g 202 t  
ORIGIN  
Query Match 1.4%; Score 21; DB 13; Length 1022;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 AACAGGCCCTGGCGGTGGC 1132  
|||||  
Db 573 AACAGGCCCTGGCGGTGGC 593.

Search completed: April 8, 2003, 23:14:59  
Job time : 2274 secs

GenCore version 5.1.4\_p5.4578  
Copyright(c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 21:20:34 ; Search time 2813 Seconds  
(without alignments)  
2214.007 Million cell updates/sec

Title: US-09-595-947C-10

Perfect score: 1127

Sequence: 1 MTPQSGATVQVTRTERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q/cgn2.1/USPTO.spool/US09595947/runat\_07042003\_160404\_20061/app\_query.fasta\_1.391  
-DB=GenEmbl -OFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09595947 -CGN\_1\_1\_2496\_@runat\_07042003\_160404\_20061 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1127	100.0	1330	9 HSA133776	AJ133776 Homo sapi
2	1105	98.0	5340	9 AF234829	AF234829 Homo sapi
3	1099	97.5	165110	9 AL450311	AL450311 Human DNA
4	1089	96.6	173341	2 AC021954	AC021954 Homo sapi
5	849	75.3	1491	6 A91167	A91167 Sequence 1
6	849	75.3	1491	10 RNRELAXT	Y10619 R.norvegicu
7	819	72.7	861	10 MMU76208	U76208 Mus musculu
8	819	72.7	1861	10 AF364300	AF364300 Mus muscu
9	819	72.7	5567	10 MMTATH4B	Y09167 M.musculus
10	819	72.7	138070	2 AC109783	AC109783 Mus muscu
11	819	72.7	185806	2 AC127417	AC127417 Mus muscu
12	523	46.4	170896	2 AC011010	AC011010 Homo sapi
13	388.5	34.5	790	5 GGA012659	AJ012659 Gallus ga
14	388.5	34.5	1880	5 AF303000	AF303000 Gallus ga
15	387.5	34.4	105572	2 AC112007	AC112007 Rattus no
16	387.5	34.2	1074	5 GGA012660	AJ012660 Gallus ga
17	382	33.9	1299	5 AF123883	AF123883 Gallus ga
18	381.5	33.9	770	5 AF123884	AF123884 Gallus ga
19	380.5	33.8	1341	5 AF109014	AF109014 Gallus ga
20	378.5	33.6	1268	6 AR023709	AR023709 Sequence
21	378.5	33.6	1675	9 BC008687	BC008687 Homo sapi
22	378.5	33.6	1717	9 BC028226	BC028226 Homo sapi
23	378.5	33.6	79970	2 AC011428	AC011428 Homo sapi
24	378.5	33.6	134506	9 AC005738	AC005738 Homo sapi
25	376.5	33.4	1268	9 HSU63842	U63842 Human neuro
26	376.5	33.3	1385	10 MMU76207	U76207 Mus musculu
27	375.5	33.3	10393	10 AF303001	AF303001 Mus muscu
28	370.5	32.9	1412	10 MMTATH4A	Y07621 M.musculus
29	370.5	32.9	151044	2 AC011432	AC011432 Homo sapi
30	370	32.8	1527	10 RN067777	U67777 Rattus norv
31	370	32.8	166677	2 AC127817	AC127817 Rattus no
32	368.5	32.7	123855	2 AC102600	AC102600 Mus muscu
33	365	32.4	932	10 MMTATH4C	Y09166 M.musculus
34	365	32.4	1315	10 MMU63841	U63841 Mus musculu
35	365	32.4	1333	6 AR023715	AR023715 Sequence
36	365	32.4	174688	2 AC124395	AC124395 Mus muscu
37	359.5	31.9	735	10 MMU67776	U67776 Mus musculu
38	359.5	31.9	71538	2 AC118243	AC118243 Mus muscu
39	356.5	31.6	1719	5 AB065284	AB065284 Cynops py
40	347.5	30.8	91531	2 AC079846_3	Continuation (4 of
41	347.5	30.8	179697	9 AC023886	AC023886 Homo sapi
42	346	30.7	6123	9 AF303002	AF303002 Homo sapi
43	329.5	29.2	1176	3 AF271788	AF271788 Branchios
44	327	29.0	750	5 DRU94588	U94588 Danio rerio
45	327	29.0	1098	5 AF024535	AF024535 Danio rer

# ALIGNMENTS

RESULT 1

```

HSA133776
LOCUS HSA133776 1330 bp DNA linear PRI 19-JUN-1999
DEFINITION Homo sapiens gene for neurogenin 3.
ACCESSION AJ133776
VERSION AJ133776.1 GI:5123782
KEYWORDS bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Ravassard,P., Icard-Liepkains,C., Wiard,L., Julien,J.P. and Mallet,J.
TITLE The human neurogenin 3 homolog maps to chromosome 10q21.3 and its expression pattern is identical to that of its murine counterparts
JOURNAL Unpublished
AUTHORS Ravassard,P.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMRC 9923, Hopital de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013 PARIS, FRANCE
FEATURES
source Location/Qualifiers
1..1330
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1..1330
/feature="ngn3"
join(<1..157,321..>1330)
mRNA 1..157
/feature="ngn3"
exon 158..320
/feature="ngn3"
intron 321..1330
/feature="ngn3"
exon 322..966
/feature="ngn3"
CDS 322..966
/function="early neurogenesis"
/codon_start=1
/product="neurogenin 3"
/protein_id="CAB45384.1"
/db_xref="GI:5123783"
translation="MTPOPSGAPTVOVTRERSPPRASEDEVCTPSAPSPPTPG
NCAEBSGGCGAPKRLARGGRRPKSELASKQRRSRKKANDRRNRHDLNSA
LDALGVLPFPDDAKLKIETLRFNHYIWAALTOTLRADHSILVLEPPHPCGELG
SPGPGPGWGLSYSPVQAGSLSPAASLEERPLGLIGATSSACLSPLGSLAFSDFL"
BASE COUNT 230 a 459 c 413 g 228 t
ORIGIN
Alignment Scores:
Pred. No.: 2,26e-46 Length: 1330
Score: 1127.00 Matches: 214
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-595-947C-10 (1-214) x HSA133776 (1-1330)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 322 ATGACCCCTCAACCTTCGGGTGGCCCACTGTCCAAAGTGACCGCGGAGCGGCGTCC 381
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 382 TTCCCCAGAGCTCGGAACAGAGTACCTGCCACACGTCGCCCGCCGAGCCCACT 441
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgTlys 60

```

LDALRGVLPFPDDAKLTKIETLRFHNYIWTQTLRIADHSLYALEPAPHCGLG  
 SPQSGPDGSLYSPVSOAGSLSPAASLERPGLIGATSSACLSPGSLAFSDFL"  
 BASE COUNT 1215 a 1508 c 1514 g 1111 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 8.92e-45 Length: 5340  
 Score: 1105.00 Matches: 211  
 Percent Similarity: 99.07% Conservative: 1  
 Best Local Similarity: 98.60% Mismatches: 2  
 Query Match: 98.05% Indels: 0  
 DB: 9 Gaps: 0

US-09-595-947c-10 (1-214) x AF234829 (1-5340)

Qy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 Db 3022 ATGACGGCTCAACCTCGGGTGCCTCCACTGTCCAAAGTGACCCGTGAGACGGAGGTCC 3081  
 Qy 21 PheProAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 Db 3082 TTCCCCAGAGCCTCGGAAGAGGAAGTACCTGCCCTCCCGCCCGCCAGCCCACT 3141  
 Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60  
 Db 3142 CGCACCGGGGAACTCGCCAGAGCGGAGAGGAGGCTCCCGAGGGGCCCCAGGAG 3201  
 Qy 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 Db 3202 CTCGGGCACGGCGGGGACGACGCGGCTTAAGACGAGTGGCACTGAGCAAGCAG 3261  
 Qy 81 ArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 Db 3262 CGACGGAGTCGGCGAAGAGGCGCAACGACGCGGCGCAATCGAATGCACAACTCAAC 3321  
 Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 Db 3322 TCGCAGCTGGAGCGCTCGCGGGTGTCTGCCACCTTCCAGACGAGCGGAGCTCACC 3381  
 Qy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 Db 3382 AAGATCGAGAGCGTGCCTTCGCCACAACTACATCTGGGCGCTGACTCAAAACGCTGGC 3441  
 Qy 141 IleAlaAspHisSerLeuThrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
 Db 3442 ATAGCGGACCACTGTACGCGCTGGAGCGCGCGCGCGCTGCGGGGAGCTGGGC 3501  
 Qy 161 SerProGlyGlyProProGlyAspTrpGlySerLeuThrSerProValSerGlnAlaGly 180  
 Db 3502 AGCCAGCGGTTCCTCCCGGGAGCTGGGGTCCCTCTACTCCAGTCTCCAGGCTGGC 3561  
 Qy 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200  
 Db 3562 AGCCTGAGTCCCGCGCTCGCTGGAGAGGAGCGCGGCTGCTGGGGGCGCCACTCTTCC 3621  
 Qy 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 Db 3622 GCCTGCTGTAGCCAGGCGAGTCTGCTCTTCTCAGATTTCG 3663

RESULT 3  
 AL450311/c 165110 bp DNA linear PRI 12-JUL-2001  
 LOCUS Human DNA sequence from clone RP11-343J3 on chromosome 10, complete  
 DEFINITION sequence.  
 ACCESSION AL450311  
 VERSION AL450311.11 GI:14626972  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 165110)  
 AUTHORS Howden, P.

# TITLE JOURNAL COMMENT

Direct Submission  
 Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Jul 8, 2001 this sequence version replaced gi:14575291.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em: EMBL; SW:  
 SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 10, constructed by the Sanger Centre Chromosome 10  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr10  
 RP11-343J3 is from the library RP11-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-343J3 The true  
 left end of clone RP11-242G20 is at 139955 in this sequence. The  
 true right end of clone RP11-404C6 is at 6588 in this sequence.

FEATURES  
 Location/Qualifiers  
 1..165110  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-343J3"  
 /clone\_lib="RP11-11.2"  
 7..147  
 /note="THE1C repeat: matches 2..142 of consensus"  
 151..293  
 /note="MIR repeat: matches 2..148 of consensus"  
 1136..1187  
 /note="26 copies 2 mer gt 98% conserved"  
 2960..3272  
 /note="AluSq repeat: matches 1..310 of consensus"  
 4144..4270  
 /note="AluSx repeat: matches 1..134 of consensus"  
 4319..4494  
 /note="AluSx repeat: matches 118..293 of consensus"  
 5662..6438  
 /note="L1MEC repeat: matches 272..1095 of consensus"  
 6916..8080  
 /note="L1MEC repeat: matches 1168..2367 of consensus"  
 9934..10087  
 /note="MIR repeat: matches 25..184 of consensus"  
 10353..10481  
 /note="MIR repeat: matches 65..194 of consensus"  
 11077..11372  
 /note="AluSx repeat: matches 1..294 of consensus"  
 11600..11639  
 /note="20 copies 2 mer tc 95% conserved"  
 12479..12608  
 /note="65 copies 2 mer at 90% conserved"  
 13433..13876  
 /note="MLTLC repeat: matches 1..466 of consensus"  
 14372..14576  
 /note="AluSq repeat: matches 1..301 of consensus"  
 14577..14633  
 /note="AluSq repeat: matches 243..299 of consensus"

```

repeat_region 14843..14943
/note="MER86 repeat: matches 9..109 of consensus"
repeat_region 1504..15613
/note="MIR repeat: matches 3..213 of consensus"
repeat_region 16626..16705
/note="MIR repeat: matches 60..140 of consensus"
repeat_region 16928..16993
/note="33 copies 2 mer gg 66% conserved"
repeat_region 17476..17569
/note="MER81 repeat: matches 2..114 of consensus"
repeat_region 17719..18069
/note="LMC4 repeat: matches 7617..7977 of consensus"
repeat_region 18117..18345
/note="MIR repeat: matches 7..262 of consensus"
repeat_region 19169..19400
/note="MIR repeat: matches 5..239 of consensus"
repeat_region 19436..19627
/note="LMC5 repeat: matches 7728..7917 of consensus"
repeat_region 19642..19716
/note="MIR repeat: matches 48..131 of consensus"
repeat_region 19969..20010
/note="21 copies 2 mer tg 100% conserved"
repeat_region 20623..20704
/note="41 copies 2 mer gt 85% conserved"
repeat_region 20738..20773
/note="U2 repeat: matches 1..36 of consensus"
repeat_region 20802..20863
/note="MLT1J repeat: matches 1..62 of consensus"
repeat_region 20889..21263
/note="MLT1F repeat: matches 188..541 of consensus"
repeat_region 21463..21618
/note="MIR repeat: matches 46..192 of consensus"
repeat_region 22019..22326
/note="AluJb repeat: matches 1..306 of consensus"
repeat_region 22381..22564
/note="L2 repeat: matches 2453..2629 of consensus"
repeat_region 22896..23174
/note="MLT1J repeat: matches 117..413 of consensus"
repeat_region 23215..23346
/note="MIR repeat: matches 48..188 of consensus"
repeat_region 23388..23532
/note="L2 repeat: matches 2097..2230 of consensus"
repeat_region 23533..23837
/note="AluJb repeat: matches 1..303 of consensus"
repeat_region 23838..24137
/note="L2 repeat: matches 1754..2097 of consensus"
repeat_region 24291..24581
/note="AluX repeat: matches 1..300 of consensus"
repeat_region 24653..24850
/note="MIR repeat: matches 1..200 of consensus"
misc_feature 26189..26685
/note="CpG island"
/evidence=not_experimental
misc_feature 29453..30918
/note="CpG island"
/evidence=not_experimental
repeat_region 31084..31156
/note="MIR repeat: matches 65..138 of consensus"
repeat_region 31618..31876
/note="AluJo repeat: matches 29..275 of consensus"
repeat_region 32767..32830
/note="MIR repeat: matches 76..139 of consensus"
repeat_region 33050..33178
/note="43 copies 3 mer tcc 72% conserved"
misc_feature 35112..36201
/note="CpG island"
/evidence=not_experimental
repeat_region 36551..36604
/note="27 copies 2 mer ac 94% conserved"
repeat_region 37585..38254
/note="LIPAL5 repeat: matches 5480..6157 of consensus"
repeat_region 38265..38561
/note="AluX repeat: matches 1..300 of consensus"
repeat_region 39214..39423
/note="MIR repeat: matches 3..219 of consensus"
repeat_region 39489..39801
/note="AluX repeat: matches 1..292 of consensus"
repeat_region 40050..40189
/note="MIR repeat: matches 1..144 of consensus"
repeat_region 41057..41290
/note="MIR repeat: matches 8..240 of consensus"
repeat_region 41296..41460
/note="L2 repeat: matches 2569..2730 of consensus"
repeat_region 41506..41944
/note="L2 repeat: matches 1916..2416 of consensus"
repeat_region 42388..42698
/note="L2 repeat: matches 1448..1779 of consensus"
repeat_region 44193..44579
/note="THE1C repeat: matches 1..371 of consensus"
repeat_region 44600..44956
/note="LTR16A repeat: matches 90..445 of consensus"
repeat_region 45240..45300
/note="MER58A repeat: matches 37..97 of consensus"
repeat_region 45798..45909
/note="LTR41 repeat: matches 90..192 of consensus"
repeat_region 46826..46871
/note="23 copies 2 mer gt 97% conserved"
repeat_region 50584..50969
/note="AluSq repeat: matches 1..287 of consensus"
repeat_region 50980..51291
/note="AluX repeat: matches 1..312 of consensus"
repeat_region 52222..52519
/note="AluX repeat: matches 1..300 of consensus"
repeat_region 54065..54260
/note="LM4 repeat: matches 3865..4055 of consensus"
repeat_region 54261..54432
/note="FAM repeat: matches 2..167 of consensus"
repeat_region 54433..54629
/note="LM4 repeat: matches 3652..3865 of consensus"
repeat_region 54648..54862
/note="LTR41 repeat: matches 11..217 of consensus"
repeat_region 54863..55236
/note="MLT1A1 repeat: matches 1..365 of consensus"
repeat_region 55237..55700
/note="LTR41 repeat: matches 217..716 of consensus"

```

Alignment Scores:  
Pred. No.: 3,72e-43 Length: 165110  
Score: 1099.00 Matches: 210  
Percent Similarity: 98.60% Conservative: 1  
Best Local Similarity: 98.13% Mismatches: 3  
Query Match: 97.52% Indels: 0  
DB: 9 Gaps: 0

US-09-595-947C-10 (1-214) x AL450311 (1-165110)

Qy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
|||||  
Db 30325 ATGACGGCTCAACCTCGGGTGGCGCCACTGTCCAGTGACCGGTGACGAGGGGGTCC 30266  
|||||

Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
|||||  
Db 30265 TTCCCCAGAGCTCGGAAGACGAAGTGCCTGCCACGTGCCCGCCGCCAGCCCACT 30206  
|||||

Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyCysArgGlyAlaProArgLys 60  
|||||  
Db 30205 CGCACACGGGGAACTCGCAGAGCGGAGGAGGAGGCTCCGAGGGGGCCCCGAGGAAG 30146  
|||||

Qy 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
|||||  
Db 30145 CTCGGGCACGGCGCGGGGAGCGGACGCGCGGCTAAGAGCGAGTTGGCACTGAGCAACGAG 30086  
|||||

Qy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
|||||  
Db 30085 CGACGGAGTCGGCGAAAGAGGCCAACACCGCGAGCGCAATCGAATGCACAACTCAAC 30026  
|||||

QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 |||||  
 Db 30025 TGGGACTGGAGCGCTGGCGTGTCTGCCACCTTCCAGACGACGCGAAGCTCAC 29966  
 |||||  
 QY 121 LysIleGluThrLeuArgPheAlaHisAanTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 |||||  
 Db 29965 AAGATCGAGACGCTGGCTTCCGCCACAACTACATCTGGGGCTGACTCAACAGCTGGCC 29906  
 |||||  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
 |||||  
 Db 29905 ATAGCGGACACAGCTGTACCGCTGGAGCGCGCGCCGACCTGCGGGAGCTGGGC 29846  
 |||||  
 QY 161 SerProGlyGlyProProGlySerPrpGlySerLeuTyrSerProValSerGlnAlaGly 180  
 |||||  
 Db 29845 AGCCAGGCGGTTCGCCCGGAGCTGGGGTCCCTCTACTCCAGCTCTCCAGGCTGGC 29786  
 |||||  
 QY 181 SerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSerSer 200  
 |||||  
 Db 29785 AGCTGAGTCCGCGCGCTGCTGGAGGAGCGACCGGGCTGCTGGGGCCACCTTTCC 29726  
 |||||  
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 |||||  
 Db 29725 GCCTGTTGAGCCGAGCTGCTGGCTTTCTCAGATTTCGT 29684  
 |||||

RESULT 4  
 AC021954/c  
 LOCUS AC021954 Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT  
 DEFINITION SEQUENCE, 24 unordered pieces.  
 ACCESSION AC021954  
 VERSION AC021954.3 GI:7417809  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 173341)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
 DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lander,T., Lehoczeky,J., Levine,R., Liu,C., Liu,G., Locke,K.,  
 Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
 McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 173341)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Lander,T., Lehoczeky,J.,  
 Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Roqov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 5, 2000 this sequence version replaced gi:6984451.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L5931  
 Center clone name: 57\_E12  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 161190 bases at least Q40  
 Consensus quality: 166837 bases at least Q30  
 Consensus quality: 168995 bases at least Q20  
 Insert size: 176000; agarose-1p  
 Insert size: 171041; sum-of-contigs  
 Quality coverage: 3.7 in Q20 bases; agarose-1p  
 Quality coverage: 3.8 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 24 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1574 1674: contig of 1574 bp in length  
 \* 1575 1674: gap of 100 bp  
 \* 1675 3043: contig of 1369 bp in length  
 \* 3044 3143: gap of 100 bp  
 \* 3144 5889: contig of 2746 bp in length  
 \* 5890 5989: gap of 100 bp  
 \* 5990 8979: contig of 2990 bp in length  
 \* 8980 9079: gap of 100 bp  
 \* 9080 13674: contig of 4595 bp in length  
 \* 13675 13774: gap of 100 bp  
 \* 13775 18831: contig of 5057 bp in length  
 \* 18832 18931: gap of 100 bp  
 \* 18932 23526: contig of 4595 bp in length  
 \* 23527 23626: gap of 100 bp  
 \* 23627 27386: contig of 3760 bp in length  
 \* 27387 27486: gap of 100 bp  
 \* 27487 32572: contig of 5086 bp in length  
 \* 32573 32672: gap of 100 bp  
 \* 32673 38632: contig of 5960 bp in length  
 \* 38633 38732: gap of 100 bp  
 \* 38733 43735: contig of 5003 bp in length  
 \* 43736 43835: gap of 100 bp  
 \* 43836 49020: contig of 5185 bp in length  
 \* 49021 49120: gap of 100 bp  
 \* 49121 53660: contig of 4540 bp in length  
 \* 53661 53760: gap of 100 bp  
 \* 53761 59544: contig of 5784 bp in length  
 \* 59545 59644: gap of 100 bp

TITLE  
 JOURNAL  
 COMMENT

\* 59645 66395: contig of 6751 bp in length  
 \* 66396 66495: gap of 100 bp  
 \* 66496 74696: contig of 8201 bp in length  
 \* 74697 74796: gap of 100 bp  
 \* 74797 81326: contig of 6530 bp in length  
 \* 81327 81426: gap of 100 bp  
 \* 81427 89446: contig of 8020 bp in length  
 \* 89447 89546: gap of 100 bp  
 \* 89547 96458: contig of 6913 bp in length  
 \* 96460 96559: gap of 100 bp  
 \* 96560 106498: contig of 9939 bp in length  
 \* 106499 106598: gap of 100 bp  
 \* 106599 116986: contig of 10388 bp in length  
 \* 116987 117086: gap of 100 bp  
 \* 117087 128890: contig of 11804 bp in length  
 \* 128891 128990: gap of 100 bp  
 \* 128991 147290: contig of 18300 bp in length  
 \* 147291 147390: gap of 100 bp  
 \* 147391 173341: contig of 25951 bp in length.

## FEATURES

source

Location/Qualifiers  
 1..173341  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /map="10"  
 /clone="RP11-57E12"  
 /clone\_lib="RPC1-11 Human Male BAC"

1..1574

/note="assembly\_fragment"

1675..3043

/note="assembly\_fragment"

3144..5889

/note="assembly\_fragment"

5990..8979

/note="assembly\_fragment"

9080..13674

/note="assembly\_fragment"

13775..18831

/note="assembly\_fragment"

18932..23526

/note="assembly\_fragment"

23627..27386

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left

27487..32572

/note="assembly\_fragment"

32673..38632

/note="assembly\_fragment"

38733..43735

/note="assembly\_fragment"

43836..49020

/note="assembly\_fragment"

49121..53660

/note="assembly\_fragment"

53761..59544

/note="assembly\_fragment"

59645..66395

/note="assembly\_fragment"

66496..74696

/note="assembly\_fragment"

74797..81326

/note="assembly\_fragment"

81427..89446

/note="assembly\_fragment"

89547..96459

/note="assembly\_fragment"

96560..106498

/note="assembly\_fragment"

106599..116986

/note="assembly\_fragment"

117087..128890

/note="assembly\_fragment"

/note="assembly\_fragment"

clone\_end:T7  
 vector\_side:right  
 128991..147290  
 /note="assembly\_fragment"  
 147391..173341  
 /note="assembly\_fragment"  
 BASE COUNT 44173 a 40928 c 40432 g 45501 t 2307 others

## Alignment Scores:

Pred. No.: 1.17e-42 Length: 173341  
 Score: 1089.00 Matches: 209  
 Percent Similarity: 98.13% Conservative: 1  
 Best Local Similarity: 97.66% Mismatches: 4  
 Query Match: 96.63% Indels: 0  
 DB: 2 Gaps: 0

US-09-595-947C-10 (1-214) x AC021954 (1-173341)

QY 1 MethProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20

Db 6765 ATGACGGCTCAACCTCGGGTGCCTCCACTGTCCAAGTGACCGTGAGCGAGCGGTCC 6706

QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40

Db 6705 TTCCCCAGAGCTCGGAAGACGAAGTGACCTGCCACGTCGCCGCCGCCGCCGCCACT 6646

QY 41 ArgThrProGlyAsnCysAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60

Db 6645 CGCACACGGGGAACTGCGCAGAGCGGAGAGGAGGCTGCGAGGGGCCCGGAGGAAG 6586

QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80

Db 6585 CTCGGGCACGGCGGGGAGCGGCGGCTTAAGACGAGTGGCTGAGCAAGCAG 6526

QY 81 ArgArgSerArgArgLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100

Db 6525 CGACGGAGTCGGCGAAAGAGCAACGACGCGAGCGCAATCGAATCAACAACCTCAAC 6466

QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120

Db 6465 TCGGCACCTGGACGCCCTGCGCGGTCTCTGCCACCTTCCAGAGCGAGCGAAGCTCAC 6406

QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTTPAlaLeuThrGlnThrLeuArg 140

Db 6405 AAGATCGAGAGCGTGGCTTCGCCCAACATCTCGGGCGTGTCTCAACGCTGGCG 6346

QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160

Db 6345 ATAGCGGACCAACAGCTTGTACGCGCTGGAGCGCGCGCGCGCACTGCGGGAGCTGGC 6286

QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnLagly 180

Db 6285 AGCCAGCGCGTTCCTCCCGGGGACTGGGGTCCCTCTACTCCCGAGTCTCCAGGCTGGC 6226

QY 181 SerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSerSer 200

Db 6225 AGCTGTAGTCCCGCGGCTGCTGGAGGAGCGCGCGGCTGCTGGGGGCCACCTTTTCC 6166

QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214

Db 6165 GCCTGCTTGAGCCCGGAGCGTCTGCTTCTCAGATTTCTG 6124

## RESULT 5

A91167

LOCUS

DEFINITION Sequence 1 from Patent WO9827206.

ACCESSION A91167

VERSION A91167.1 GI:6740202

KEYWORDS

SOURCE

ORGANISM

Rattus sp.

Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

linear

PAT 22-JAN-2000



```

Rattus
1 (bases 1 to 1491)
Icard-Liepkalns,C., Mallet,J. and Corresponding,N.A.
Patent: WO 9827406-A 1 25-JUN-1998;
JOURNAL ICARD LIEPKALNS CHRISTINE (FR); MALLETT JACQUES (FR)
FEATURES
Source
Location/Qualifiers
1..1491
/organism="Rattus sp."
/db_xref="taxon:10118"
/feature="unnamed protein product"
/protein_id="CAB69414.1"
/db_xref="GI:6740203"
/translation="MAPHLDAPTIQVSOEQPPGASDHEVLSNSTPPSPTLVPR
DCSEAAGDCRGTRKURARRPKSELALSKORRRKKANDRENRHNLNSA
LDALRGVLPDPDAKLTETFAHNYIWAOTLRADHSFYGPVPPVPCGELGS
PGGSSGDWGSYSPVSOAGSLPTASLEEFPLQVPSPSCLLPGLTVFSDFL"
BASE COUNT 307 a 487 c 413 g 284 t
ORIGIN
Alignment Scores:
Pred. No.: 5,67e-33 Length: 1491
Score: 849.00 Matches: 166
Percent Similarity: 83.72% Conservative: 14
Best Local Similarity: 77.21% Mismatches: 33
Query Match: 75.33% Indels: 2
DB: 6 Gaps: 2
US-09-595-947C-10 (1-214) x A91167 (1-1491)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGGCGCCTCATCCCTTGGATCGCCACCATCAAGTCCCAAGTCCCAAGAGACCCAGCAACCC 518
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTTCCCGGAGCCTCGGACCAAGTGTCTCAGTTCCTCAATCCACCACCCAGTACGCCACT 578
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGTACCGAGGAGTGTCTCCGAAGCAGAGCAGTGTGCTGCGGAGGACATCGAGGAAG 638
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCGGTGGCGGCGGAGGCGCACAGCCCAAGAGCGAGTGTGCTGAGCAAGCAG 698
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100
DB 699 CGACGAGCGCGGCAAGAGCCACAGCAGCGGAGGACCAAGCATGCGACACCTTAC 758
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
DB 759 TCCGGCTGGATGGCTGCGGCTGTCTGCGCACCTTCCCGATCGACGCAACTTACA 818
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrArg 140
DB 819 AGATCGAGACCTCGCTTCCGCCCAACTACATTGGGCACCTGACTCAGACGCTGCC 878
QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaPheHisCysGlyGluLeuGly 160
DB 879 ATAGCGGACCAAGCTTCTACGCGCCCGAGCCCTTGTGCC---TGTGGGAGCTGGA 935
QY 161 SerPro---GlyClyProGlyAspThrPglySerLeuThrSerProValSerGlnAla 179
DB 936 AGCCGGGAGGGGCTCCAGCGGAGTGGGCTCTATCTACTCCCAAGT 995
QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuAlaThrSer 199
DB 996 GGTAGCTGAGGCCACAGCCTCATTTGGAGGATTCCTTGGCTGCGAGTGGCCAGCTCC 1055
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 1055

```

```

Db 1056 CCATCTGTCTGCTCCGGGACCCCTGGTGTCTCAGACTTCTTG 1100
RESULT 6
RNRELAXT
LOCUS R.norvegicus mRNA for transcriptional regulator, Relax.
DEFINITION Y10619
ACCESSION Y10619
VERSION Y10619.1 GI:2072737
KEYWORDS Relax; transcriptional regulator.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1491)
AUTHORS Ravassard,P., Chatail,F., Mallet,J. and Icard-liepkalns,C.
TITLE Relax, a novel rat bHLH transcriptional regulator transiently
expressed in the ventricular proliferating zone of the developing
central nervous system
J. Neurosci. Res. 48 (2), 146-158 (1997)
JOURNAL 97276390
MEDLINE 9130143
REFERENCE 2 (bases 1 to 1491)
AUTHORS Ravassard,P.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,
Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013
Paris, FRANCE
FEATURES
Location/Qualifiers
1..1491
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/clone_lib="RT-PCR"
459..1103
/function="transcriptional regulator"
/codon_start=1
/product="Relax"
/protein_id="CAA71630.1"
/db_xref="GI:2072738"
/translation="MAPHLDAPTIQVSOEQPPGASDHEVLSNSTPPSPTLVPR
DCSEAAGDCRGTRKURARRPKSELALSKORRRKKANDRENRHNLNSA
LDALRGVLPDPDAKLTETFAHNYIWAOTLRADHSFYGPVPPVPCGELGS
PGGSSGDWGSYSPVSOAGSLPTASLEEFPLQVPSPSCLLPGLTVFSDFL"
BASE COUNT 307 a 487 c 413 g 284 t
ORIGIN
Alignment Scores:
Pred. No.: 5,67e-33 Length: 1491
Score: 849.00 Matches: 166
Percent Similarity: 83.72% Conservative: 14
Best Local Similarity: 77.21% Mismatches: 33
Query Match: 75.33% Indels: 2
DB: 10 Gaps: 2
US-09-595-947C-10 (1-214) x RNRELAXT (1-1491)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGGCGCCTCATCCCTTGGATCGCCACCATCAAGTCCCAAGTCCCAAGAGACCCAGCAACCC 518
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTTCCCGGAGCCTCGGACCAAGTGTCTCAGTTCCTCAATCCACCACCCAGTACGCCACT 578
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGTACCGAGGAGTGTCTCCGAAGCAGAGCAGTGTGCTGCGGAGGACATCGAGGAAG 638
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCGGTGGCGGCGGAGGCGCACAGCCCAAGAGCGAGTGTGCTGAGCAAGCAG 698

```

```

QY 81 ArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
    |||||
Db 699 CGACGAAGCCGGCCAGAGAGCAACGACGGAGCGCAACCGCATGCACACCTTAC 758
    |||||
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
    |||||
Db 759 TCCGCGCTGGATGCGTGCAGGCTGCTGCCACCTTCCGGATGACGCGCAAACTTACA 818
    |||||
QY 121 LysileGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
    |||||
Db 819 AAGATCAGACCTTCCCTTCGCCCACTACATATGGGACAGTACAGACCTCGCG 878
    |||||
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisLysCysGlyGluLeuGly 160
    |||||
Db 879 ATAGCGGACACAGCTTCTACGGCCCGAGCCCTCTGCC---TGTGGGAGCTGGGA 935
    |||||
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
    |||||
Db 936 AGCCCGGAGGGGCTCCAGCGGCTGCTTCTACTCTCTACTCTCTCTCTCTCTCTCT 995
    |||||
QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSer 199
    |||||
Db 996 GGTAGCTGAGCCCAAGCTCTATGAGAGAGTTCCTCGCCGCGAGGCGCCAGCTCC 1055
    |||||
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
    |||||
Db 1056 CCATCCTGTCTGCTCCCGGACCCCTGGTGTCTCTCAGACTTCITG 1100
    |||||

RESULT 7
MMU76208
LOCUS Mus musculus neurogenin 3 (ngn3) gene, complete cds.
DEFINITION
ACCESSION U76208
VERSION U76208.1 GI:1815654
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Sommer, L., Ma, Q. and Anderson, D.J.
TITLE neurogenins, a novel family of atonal-related bHLH transcription
factors, are putative mammalian neuronal determination genes that
reveal progenitor cell heterogeneity in the developing CNS and PNS
Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
JOURNAL
MEDLINE 97153565
PUBMED 9000438
REFERENCE
AUTHORS Sommer, L., Ma, Q. and Anderson, D.J.
TITLE Direct Submission
SUBMITTED (24-OCT-1996) Biology 216-76, California Institute of
Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
USA
COMMENT On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES
    source
        1..861
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /protein_id="AAC53029.1"
            /db_xref="GI:1666912"
            /gene="ngn3"
            /gene="ngn3"
            /gene="ngn3"
            /function="bHLH transcription factor"
            /note="NGN3"
            /codon_start=1
            /product="neurogenin 3"
            /protein_id="AAC53029.1"
            /db_xref="GI:1666912"
            /translation="WAPHPDLALTIQVSPETQOPFGASDHEVLSSNTPPTLIPR
            DCSEAVGDCRTSKRLARRGRNRPKSELASKORRRKANDRERNRNLNSA
            LDALRGVLPFPDDAKLTKIETLRFAHNYIWTOTLRADHSFVGPEPPVCGEUGS
            PGGSGNDWGSIIYSPVQAGNLSPTASLEEFPLQLVPSSPSYLLPGLVFSDFL"

```

```

BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN
Alignment Scores: 9.6e-32 Length: 861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 10 Gaps: 2
US-09-595-947c-10 (1-214) x MMU76208 (1-861)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
    |||||
Db 160 ATGGCGCTCATCCTTGGATGCGCTCACCATCCAGTGTCCCGAGAGACACAACT 219
    |||||
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
    |||||
Db 220 TTTCCCGGAGCTCGGACACGAAAGTCTCAGTTCCAAATCCACCACCTAGCCCACT 279
    |||||
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
    |||||
Db 280 CTCATCCTAGGGACTCTCCGAGCAGAAAGTGGTGACTGCCGAGGGAGCTCCGAGGA 339
    |||||
QY 61 LeuArgAlaArgArgGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
    |||||
Db 340 CTCCGCGCCGACGCGGAGGGGCGCAACAGGCCCCAAGAGGAGTGGCACTCAGCAAC 399
    |||||
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
    |||||
Db 400 CGAAGAAGCCGGCCAGAGAGGCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 459
    |||||
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
    |||||
Db 460 TCGCGCTGGATGCGCTGCGCGGTCTCTCCACCTCCCGGATGACGCCCAACTTACA 519
    |||||
QY 121 LysileGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
    |||||
Db 520 AAGATCAGACCTTCCCTTCGCCCACTACATCTGGGACATGACTCAGACCTCGCG 579
    |||||
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
    |||||
Db 580 ATAGCGGACACAGCTTCTATGCGCGGAGCCCTCTGCC---TGTGGAGAGCTGGG 636
    |||||
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
    |||||
Db 637 AGCCCGGAGGTGGCTCCAGGGGAGTGGGCTCTATCTACTCCCGAGTCTCCCAAGCG 696
    |||||
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
    |||||
Db 697 GGTAACTGAGCCCAAGGCTCATTTGAGGAATTCCTGCCCTGAGGTGCCCAAGTCC 756
    |||||
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
    |||||
Db 757 CCATCCTATCTGCTCCCGGAGCAGCTGGTGTCTCAGACTCTTG 801
    |||||

RESULT 8
AF364300
LOCUS Mus musculus neurogenin 3 gene, complete cds.
DEFINITION
ACCESSION AF364300
VERSION AF364300.1 GI:13937128
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Lee, J., Smith, S., Wataha, H., Lin, J., Scheel, D., Wang, J., Mirmira, R.
TITLE Regulation of the pancreatic pro-endocrine gene neurogenin3
JOURNAL Diabetes (2001) In press

```

REFERENCE 2 (bases 1 to 1861)  
 AUTHORS Schwitzgebel,V. and German, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-MAR-2001) Hormone Research Institute, University of California San Francisco, 513 Parnassus Ave, HSW1090, San Francisco, CA 94145-0534, USA

FEATURES  
 source 1. 1861  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 <1093. .>1737  
 /product="neurogenin 3"  
 1093. .1737  
 /codon\_start=1  
 /product="neurogenin 3"  
 /protein\_id="AAK50058.1"  
 /db\_xref="GI:13937129"  
 /translation="MAPHLDALTIQVSPQQPPGASDEVLSNSTPSPPTLIPR  
 DCSAEVGDGCRGTRKRLRRGGRNPKSELALSKRRSRKRRKANDERNRMHLNSA  
 LDALRGVLPFPDDAKLTETLFAHNYIWAQTQTLRIADHSFYGPPEPPVPCGELGS  
 PGGSGNDWGSIIYSPVQAGNLSPTASLEEFPGLOVPSSPSYLLPGALVFSEFL"

BASE COUNT 397 a 560 c 537 g 367 t

ORIGIN

Alignment Scores:  
 Pred. No.: 1.91e-31 Length: 1861  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.33% Conservative: 14  
 Best Local Similarity: 75.81% Mismatches: 36  
 Query Match: 72.67% Indels: 2  
 DB: 10 Gaps: 2

US-09-595-947C-10 (1-214) x AF364300 (1-1861)

Qy 1 MethThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 Db 1093 ATGGCGCCTCATCCCTTGGATCGCTCACCATCCAAAGTTCCTCCAGAGACACAACT 1152

Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerSerAlaProSerProThr 40  
 Db 1153 TTTCGGGAGCCTCGGACGACGAAAGTGTCTCAATTCACCCCTAGCCCTAGCCCACT 1212

Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60  
 Db 1213 CTCATCTAGGACTGCTCCGACACAGAGTGGTGTACTGCCGAGGACCTCGAGGAG 1272

Qy 61 LeuArgAlaArgArgGlyCysArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 Db 1273 CTCGCGCCCGACGGGGGCGCAACAGCCCAAGAGCGAGTGGCCTCAGCAACAG 1332

Qy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100  
 Db 1333 CGAAGAGCCGCGCGCAAGAAGCCCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 1392

Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 Db 1393 TCGGCGCTGGATGCGCTCGGGGTCTCTGCCACCTTCCCGGATGAGCGCAAACTTACA 1452

Qy 121 LysLeuGluThrLeuArgPheAlaHisAsnTyrlleThrAlaLeuThrGlnThrLeuArg 140  
 Db 1453 AAGATCGAGACCTCGCTCGCCACAACTACATCTGGCCTGACTGACTGAGCGCTGCGC 1512

Qy 141 IleAlaAspHisSerLeuTyrrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 Db 1513 ATAGCGGACACAGCTTCTATGGCCCGGAGCCCTGTGCCCTGTCGCTGAGAGCTGGGG 1569

Qy 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrrSerProValSerGlnAla 179  
 Db 1570 AGCCCGGAGGTGGCTCCACGGGAGCTGGGGCTCTATCTACTCCCTAGCTCCCAAGCG 1629

Qy 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 Db 1630 GGTAACCTGAGCCCCACGGCCTCATTTGGAGGAATTCCTCGCTGCTGAGTGGCCAGCTCC 1689

Qy 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 Db 1690 CCATCCTATCTGCTCCCGGAGCACTGGTGTCTCAGACTTCTTG 1734

RESULT 9  
 MMATH4B  
 LOCUS M.musculus MATH4B gene. 5567 bp DNA linear ROD 25-OCT-2000  
 DEFINITION Y09167  
 ACCESSION Y09167.2 GI:11065737  
 VERSION BHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.  
 KEYWORDS house mouse.  
 SOURCE Mus musculus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 5567)  
 AUTHORS Cau,E., Gradwohl,G., Fode,C. and Guillemot,F.  
 TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron progenitors  
 JOURNAL Development 124 (8), 1611-1621 (1997)  
 MEDLINE 97461963  
 PUBMED 9108377  
 REFERENCE 2 (bases 1 to 5567)  
 AUTHORS Jacquemin,P., Durvieux,S.M., Jensen,J., Godfraind,C., Gradwohl,G., Guillemot,F., Madsen,O.D., Carmeliet,P., Dewerchin,M., Collen,D., Rousseau,G.G. and Lemaigre,F.P.  
 TITLE Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3  
 JOURNAL Mol. Cell. Biol. 20 (12), 4445-4454 (2000)  
 MEDLINE 20285449  
 PUBMED 10825208  
 REFERENCE 3 (bases 1 to 5567)  
 AUTHORS Gradwohl,G.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC, CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg, F-67404 ILLKIRCH cedex, FRANCE  
 REMARK Revised by [4]  
 REFERENCE 4 (bases 1 to 5567)  
 AUTHORS Lemaigre,F.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic Research Unit, Louvain University Medical School, Avenue Hippocrate 75, box 7529, Brussels 1200, BELGIUM  
 COMMENT On Oct 31, 2000 this sequence version replaced gi:1666087.  
 FEATURES  
 source 1. 5567  
 /organism="Mus musculus"  
 /strain="SV129"  
 /db\_xref="taxon:10090"  
 /clone="lambda GEM12 clone 19C"  
 /cell\_line="D3"  
 /cell\_type="embryonic stem cells"  
 /clone\_lib="Genomic library No.62, MboI partial digests cloned in BamHI"  
 /dev\_stage="embryos"  
 958  
 misc\_binding /bound\_moiety="HNF6"  
 3692  
 misc\_binding /bound\_moiety="HNF6"  
 4145. .5567  
 gene /gene="MATH4B"  
 4145. .4151  
 TATA\_signal /gene="MATH4B"  
 /note="putative"  
 4923. 5567  
 CDS /gene="MATH4B"  
 /codon\_start=1  
 /product="neurogenin 3"  
 /protein\_id="CAA70366.1"  
 /db\_xref="GI:1666088"

/db\_xref="SWISS-PROT:F70661"  
 /translation="NAPPLDALTIOVSPETQPPPGASDHVLSNSNTPPTLIPR  
 DCSAEVGCCTCTSKLRARCGRRPKSELALSORSRRKKANDRENRMHNLNSA  
 LDALRGVLPDPDDAKLKIETLRFAHNYIWAQTQLRIADSHFDPPEYPCGELGS  
 PGGSGNGWMSIYSPVSGAGNISPASUEEFGLOVSPSPSYLLPGLVALISDFL"  
 BASE COUNT 1271 a 1549 c 1564 g 1183 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,09e-31 Length: 5567  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.33% Conservative: 14  
 Best Local Similarity: 75.81% Mismatches: 36  
 Query Match: 72.67% Indels: 2  
 DB: 10 Gaps: 2

US-09-595-947c-10 (1-214) x MMATH4B (1-5567)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 4923 ATGGCGCTCATCCCTTGGATGGCTCACCATCAAGTGTCCCGCAGACACAAACCT 4982  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 4983 TTTCGCGGAGCCPCGACACGACGAGTCTCAATTCACCCACCTAGCCCACT 5042  
 QY 41 ArgThrProGlyAsnGlyAlaGluGluGlyCysArgGlyAlaProArgLys 60  
 DB 5043 CTCATACCTAGGACGTCTCCGAGGAGAGAGTGGTGACTGCGAGGGACCTCGAGGAAG 5102  
 QY 61 LeuAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 5103 CTCGCGCGCGACGCGAGGCGCAACAGCCAGAGCGAGTGGCACTCAGCAAAACAG 5162  
 QY 81 ArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 DB 5163 CGAAGAAGCGCGGCAAGAGGCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 5222  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120  
 DB 5223 TCGCGCTGGATCGCTGCGGTGTCTCCGACCTTCCGATGACGCGCAACTTACA 5282  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 DB 5283 AAGATCGAGACCTGCGCTCGCCCACTACATCTGGCACTGACTCAGACGCTGCGC 5342  
 QY 141 IleAlaAsnHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
 DB 5343 ATAGCGGACACAGCTTCTATGCGCGGAGCCCTGTGCCC---TGTGGAGAGCTGGGG 5399  
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179  
 DB 5400 AGCCCGGAGGTGGCTCCACGCGGACTGGGGCTCTATCTACTCCCACTGCCAAGCG 5459  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSer 199  
 DB 5460 GGTAACTGAGCCCAACGCGCTCATTTGGAGGAATTCCTGCGCTCGAGGTCCTCC 5519  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 DB 5520 CCATCTATCTGCTCCGCGGAGCACTGGTGTCTCAGACTCTTG 5564

RESULT 10  
 AC109783/c  
 LOCUS AC109783 138070 bp DNA linear HTG 07-FEB-2002  
 DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17  
 unorderd pieces.  
 ACCESSION AC109783  
 VERSION AC109783.1 GI:18581594  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

# REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 138070)  
 McCombie,W.R., Baker,J.P., Ballja,V., Dedhia,N.N., de la  
 Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,  
 Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,  
 Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and  
 Zutavern,T.  
 Mouse Genomic Sequence  
 Unpublished  
 2 (bases 1 to 138070)  
 McCombie,W.R.  
 Direct Submission  
 Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing  
 Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring  
 Harbor, NY 11724, USA  
 ----- Genome Center -----  
 Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor  
 Laboratory  
 Center code: CSHL  
 Web site: <http://www.cshl.org/genseq>  
 Contact: mcombie@cshl.org  
 ----- Project Information -----  
 Project name: RP23-121F10  
 Clone name: RP23-121F10  
 Insert size: 173000; agarose-fp  
 Insert size: 141616; sum-of-contigs  
 Quality coverage: 4.00 in Q20 bases; agarose-fp  
 Quality coverage: 3.70 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 17 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 17785: contig of 17785 bp in length  
 \* 17786: 17874: gap of unknown length  
 \* 17875: 30356: contig of 12482 bp in length  
 \* 30357: 30444: gap of unknown length  
 \* 30445: 42306: contig of 11862 bp in length  
 \* 42307: 42394: gap of unknown length  
 \* 42395: 53598: contig of 11204 bp in length  
 \* 53599: 53687: gap of unknown length  
 \* 53687: 64435: contig of 10669 bp in length  
 \* 64435: 64443: gap of unknown length  
 \* 64444: 74016: contig of 9573 bp in length  
 \* 74017: 83366: contig of 9262 bp in length  
 \* 83367: 83454: gap of unknown length  
 \* 83455: 92355: contig of 8901 bp in length  
 \* 92356: 92443: gap of unknown length  
 \* 92444: 100821: contig of 8378 bp in length  
 \* 100822: 100909: gap of unknown length  
 \* 100910: 107529: contig of 6620 bp in length  
 \* 107530: 107617: gap of unknown length  
 \* 107618: 114066: contig of 6449 bp in length  
 \* 114067: 114154: gap of unknown length  
 \* 114155: 118873: contig of 4719 bp in length  
 \* 118874: 118961: gap of unknown length  
 \* 118962: 123619: contig of 4658 bp in length  
 \* 123620: 123707: gap of unknown length  
 \* 123708: 128240: contig of 4533 bp in length  
 \* 128241: 128328: gap of unknown length  
 \* 128329: 132682: contig of 4354 bp in length  
 \* 132683: 132770: gap of unknown length  
 \* 132771: 136341: contig of 3571 bp in length  
 \* 136342: 136429: gap of unknown length  
 \* 136430: 138070: contig of 1641 bp in length.

Location/Qualifiers

1. 138070  
 /organism="Mus musculus"

## FEATURES

source

/db\_xref="taxon:10090"  
/clone="RP23-121F10"

BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others  
ORIGIN

Alignment Scores: 8.97e-30 Length: 138070  
Pred. No.: 819.00 Matches: 163  
Score: 82.33% Conservative: 14  
Best Local Similarity: 75.81% Mismatches: 36  
Query Match: 72.67% Indels: 2  
DB: 2 Gaps: 2

US-09-595-947c-10 (1-214) x AC109783 (1-138070)

Qy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
Db 111220 ATGGCGCCTCATCCCTTGGATGGCTCACCATCCCAAGTGTCCCGAGACACACAACT 111161  
Qy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40  
Db 111160 TTTCCGGAGCCTCGGACACCAAGTGTCTCAGTTCCAAATCCACCCCACTAGCCCACT 111101  
Qy 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60  
Db 111100 CTCATACCTAGGAGCTGCTCCGAAGCAGAAAGTGGTGATCGCGAGGACCTCGAGGAAG 111041  
Qy 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
Db 111040 CTCGCGCCCGAGCGGAGGCGCACAGCCCAAGAGCGAGTGTGCACATCAGCAACAG 110981  
Qy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
Db 110980 CGAAGAAGCGCGCGCAAGAAGCGCAATGATCGGGAGCGCAATGCGATGCAACACCTCAAC 110921  
Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120  
Db 110920 TCGGCGCTGGAGCGCTGCGGGTGTCTGCCACCTTCCGGATGACGCCCAACTACA 110861  
Qy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
Db 110860 AAGATCGAGACCTCGGCTTCCGCCACACTACATCTGGCACTGACTGAGACGCTGCGC 110801  
Qy 141 IleAlaAspHisSerLeuThrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
Db 110800 ATAGCGGACACAGCTTCTATGCGCGAGCCCTGTGCCCCTTCCGGATGACGCCCAACTACA 110744  
Qy 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179  
Db 110743 AGCCCGGAGGTGGCTCCACGGGAGTGGGGCTCTATCTACTCCCACTCCCAAGCG 110684  
Qy 180 GlySerLeuSerProAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
Db 110683 GGTAACTAGACCCACCGCCTCATTTGGAGAAATTCCTGGCTGAGGTGCCAGCTCC 110624  
Qy 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
Db 110623 CCATCTATCTGCTCCCGGAGCACTGGTGTCTCAGACTTCTG 110579

RESULT 11

AC127417 185806 bp DNA linear HTG 15-JUL-2002  
LOCUS  
DEFINITION Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT  
SEQUENCE, 36 unordered pieces.

AC127417  
ACCESSION AC127417.1 GI:21759524  
VERSION  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 185806)  
AUTHORS McPherson, J.D. and Waterston, R.H.

The sequence of Mus musculus clone

Unpublished

2 (bases 1 to 185806)

McPherson, J.D. and Waterston, R.H.

Direct Submission

Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

# COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
----- Project Information -----  
Center project name: M\_BA0459M02

----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 173518 bases at least Q30  
Consensus quality: 177276 bases at least Q30  
Consensus quality: 179611 bases at least Q20  
Insert size: 192000; agarose-fp  
Insert size: 184535; sum-of-contrigs  
Quality coverage: 3.66 in Q20 bases; agarose-fp  
Quality coverage: 3.85 in Q20 bases; sum-of-contrigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 36 contrigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contrigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1042: contig of 1042 bp in length  
\* 1043: gap of unknown length  
\* 1142: contig of 1074 bp in length  
\* 2216: gap of unknown length  
\* 2317: contig of 1522 bp in length  
\* 3938: gap of unknown length  
\* 3939: contig of 1458 bp in length  
\* 5397: gap of unknown length  
\* 5497: contig of 1378 bp in length  
\* 6874: gap of unknown length  
\* 6974: gap of unknown length  
\* 8481: contig of 1507 bp in length  
\* 8582: gap of unknown length  
\* 8582: contig of 1293 bp in length  
\* 9875: gap of unknown length  
\* 9875: contig of 1509 bp in length  
\* 11483: gap of unknown length  
\* 11484: contig of 2251 bp in length  
\* 13835: gap of unknown length  
\* 13935: contig of 2101 bp in length  
\* 16036: gap of unknown length  
\* 16136: contig of 2965 bp in length  
\* 19100: gap of unknown length  
\* 19200: contig of 2984 bp in length  
\* 22184: gap of unknown length  
\* 22285: contig of 3003 bp in length  
\* 25288: gap of unknown length  
\* 25388: contig of 3044 bp in length  
\* 28432: gap of unknown length  
\* 28532: contig of 2878 bp in length  
\* 31410: gap of unknown length  
\* 35320: contig of 3811 bp in length  
\* 35321: gap of unknown length  
\* 38808: contig of 3388 bp in length  
\* 38908: gap of unknown length  
\* 43846: contig of 4938 bp in length

```

* 43847 43946: gap of unknown length
* 43947 49171: contig of 5225 bp in length
* 49172 49971: gap of unknown length
* 49972 53157: contig of 3886 bp in length
* 53158 53257: gap of unknown length
* 53258 57731: contig of 4474 bp in length
* 57732 57831: gap of unknown length
* 57832 62660: contig of 4829 bp in length
* 62661 62760: gap of unknown length
* 62761 68243: contig of 5483 bp in length
* 68244 68343: gap of unknown length
* 68344 73047: contig of 4704 bp in length
* 73048 73147: gap of unknown length
* 73148 78565: contig of 5418 bp in length
* 78566 78665: gap of unknown length
* 78666 85726: contig of 7061 bp in length
* 85727 85826: gap of unknown length
* 85827 91985: contig of 6159 bp in length
* 91986 92086: gap of unknown length
* 92087 97985: contig of 5899 bp in length
* 97986 105338: contig of 7254 bp in length
* 105339 113368: gap of unknown length
* 113369 113468: contig of 7930 bp in length
* 113469 120753: gap of unknown length
* 120754 120853: gap of unknown length
* 120854 129342: contig of 8489 bp in length
* 129343 129443: gap of unknown length
* 129444 138998: contig of 9556 bp in length
* 138999 139098: gap of unknown length
* 139099 149579: contig of 10481 bp in length
* 149580 149679: gap of unknown length
* 149680 162343: contig of 12664 bp in length
* 162344 162444: gap of unknown length
* 162445 185806: contig of 23363 bp in length.

```

## Location/Qualifiers

```

1. 185806
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-459M2"

```

```

misc_feature 1..1042
/note="assembly_name:Contig24"
misc_feature 1143..2216
/note="assembly_name:Contig25"
misc_feature 2317..3838
/note="assembly_name:Contig26"
misc_feature 3939..5396
/note="assembly_name:Contig27"
misc_feature 5497..6874
/note="assembly_name:Contig28"
misc_feature 6975..8481
/note="assembly_name:Contig29"
misc_feature 8582..9874
/note="assembly_name:Contig30"
misc_feature 9975..11483
/note="assembly_name:Contig31"
misc_feature 11584..13834
/note="assembly_name:Contig32"
misc_feature 13935..16035
/note="assembly_name:Contig33"
misc_feature 16136..19100
/note="assembly_name:Contig34"
misc_feature 19201..22184
/note="assembly_name:Contig35"
misc_feature 22285..25287
/note="assembly_name:Contig36"
misc_feature 25388..28431
/note="assembly_name:Contig37"
misc_feature 28532..31409
/note="assembly_name:Contig38"
misc_feature 31510..35320

```

## FEATURES

## source

```

misc_feature /note="assembly_name:Contig39"
35421..38808
/note="assembly_name:Contig40"
38909..43846
/note="assembly_name:Contig41"
43947..49171
/note="assembly_name:Contig42"
49272..53157
/note="assembly_name:Contig43"
53258..57731
/note="assembly_name:Contig44"
57832..62660
/note="assembly_name:Contig45"
62761..68243
/note="assembly_name:Contig46"
68344..73047
/note="assembly_name:Contig47"
73148..78565
/note="assembly_name:Contig48"
78666..85726
/note="assembly_name:Contig49"
85827..91985
/note="assembly_name:Contig50"
92086..97984
/note="assembly_name:Contig51"
clone_end:SP6
vector_side:left
98085..105338
/note="assembly_name:Contig52"
105439..113368
/note="assembly_name:Contig53"
113469..120753
/note="assembly_name:Contig54"
120854..129342
/note="assembly_name:Contig55"
129443..138998
/note="assembly_name:Contig56"
139099..149579
/note="assembly_name:Contig57"
149680..162343
/note="assembly_name:Contig58"
162444..185806

```

## Alignment Scores:

```

Pred. No.: 1.17e-29 Length: 185806
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 2 Gaps: 2

```

US-09-595-947c-10 (1-214) x AC127417 (1-185806)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 158824 ATGGCGCTCATCTGGATGCTCCATCAAGTGTCCCGACAGACACAAACCT 158883
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 158884 TTTCGGGAGCCCTCGGACACAGAGTGCTCAGTTCCTCAATTCACCCACCTAGCCCACT 158943
QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
Db 158944 CTCATACCTAGGAGCTGCTCCGAAGCAGAAAGTGGGTGACTGCCGAGGAGCTCGAGGAAG 159003
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 159004 CTCGGCGCGGCGGAGGCGCAACAGGCCCAAGAGCGAGTGGCAGCTCAGCAACAG 159063
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 159064 CGAAGAGCGCGGCGCAAGAGGCCCAATGATCGGAGCGCAATCGCATGCACCAACCTCAAC 159123

```

```

QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 159124 TGGGCGCTGGATGGCTGCGCGGTGCTCCGCCACCTTCCCGATGACGCCAACTTACA 159183
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 159184 AAGATCGAGACCGCTGCGCTGCCACCACTACATCTGGCACTGACTCAGACGCTGCC 159243
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db 159244 ATAGGGGACACAGCTTCATGGCCGGAGCCCTGTCGCC---TGTGAGAGCTGGGG 159300
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 159301 ACCCGGAGGTGGCTCCACGGGACTGGGCTCTATCTATCTCCCATCTCTCCCAAGCG 159360
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 159361 GGTAACCTGAGCCCGACCGCCTCATGGAGGAATCCCTGGCTGCGAGTGCCGACGCTCC 159420
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 159421 CCATCCTATCTGCTCCCGGAGCACTGGTGTCTTCAGACTTCTTG 159465

RESULT 12
AC011010/c
LOCUS AC011010 170896 bp DNA linear HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-6P16, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
ACCESSION AC011010
VERSION AC011010.4 GI:7107881
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.

REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckhagter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
TITLE Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 28, 2000 this sequence version replaced gi:6479051.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2916
Center clone name: 6_P_16
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

Assembly program: Phrap; version 0.960731
Consensus quality: 114103 bases at least Q40
Consensus quality: 141555 bases at least Q30
Consensus quality: 158230 bases at least Q20
Insert size: 154000; agarose-fp
Insert size: 168896; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1744: contig of 1744 bp in length
* 1745 1844: gap of 100 bp
* 1845 2954: contig of 1110 bp in length
* 2955 3054: gap of 100 bp
* 3055 4839: contig of 1785 bp in length
* 4840 4939: gap of 100 bp
* 4940 7461: contig of 2522 bp in length
* 7462 7561: gap of 100 bp
* 7562 11912: contig of 4351 bp in length
* 11913 12012: gap of 100 bp
* 12013 15127: contig of 3115 bp in length
* 15128 15227: gap of 100 bp
* 15228 19176: contig of 3949 bp in length
* 19177 19276: gap of 100 bp
* 19277 24838: contig of 5562 bp in length
* 24839 24938: gap of 100 bp
* 24939 30632: contig of 5694 bp in length
* 30633 30732: gap of 100 bp
* 30733 37821: contig of 7089 bp in length
* 37822 37921: gap of 100 bp
* 37922 44638: contig of 6717 bp in length
* 44639 44738: gap of 100 bp
* 44739 49424: contig of 4686 bp in length
* 49425 49524: gap of 100 bp
* 49525 56517: contig of 6993 bp in length
* 56518 56617: gap of 100 bp
* 56618 65413: contig of 8796 bp in length
* 65414 65513: gap of 100 bp
* 65514 75659: contig of 10146 bp in length
* 75660 75759: gap of 100 bp
* 75760 86433: contig of 10674 bp in length
* 86434 86533: gap of 100 bp
* 86534 98763: contig of 12230 bp in length
* 98764 98863: gap of 100 bp
* 98864 109919: contig of 11056 bp in length
* 109920 110019: gap of 100 bp
* 110020 125264: contig of 15245 bp in length
* 125265 125364: gap of 100 bp
* 125365 141043: contig of 15679 bp in length
* 141044 141143: gap of 100 bp
* 141144 170896: contig of 29753 bp in length.
Location/Qualifiers
1..170896
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-6P16"
/clone_lib="RP11-6P16"
1..1744
/note="assembly_fragment"
1845..2954
/note="assembly_fragment"
3055..4839
/note="assembly_fragment"
4940..7461
/note="assembly_fragment"
7562..11912

```

misc\_feature /note="assembly\_fragment"  
 12013..15127  
 /note="assembly\_fragment"  
 15228..19176  
 /note="assembly\_fragment"  
 19277..24838  
 /note="assembly\_fragment"  
 24939..30632  
 /note="assembly\_fragment"  
 30733..37821  
 /note="assembly\_fragment"  
 37922..44638  
 /note="assembly\_fragment"  
 44739..49424  
 /note="assembly\_fragment"  
 49525..56517  
 /note="assembly\_fragment"  
 56618..65413  
 /note="assembly\_fragment"  
 65514..75659  
 /note="assembly\_fragment"  
 75760..86433  
 /note="assembly\_fragment"  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:left  
 86534..98763  
 /note="assembly\_fragment"  
 98864..109919  
 /note="assembly\_fragment"  
 110020..125264  
 /note="assembly\_fragment"  
 125365..141043  
 /note="assembly\_fragment"  
 141144..170896  
 /note="assembly\_fragment"  
 41547 a 41714 c 41241 g 44377 t 2017 others

Alignment Scores:  
 Pred. No.: 1.8e-15 Length: 170896  
 Score: 523.00 Matches: 99  
 Percent Similarity: 98.02% Conservations: 0  
 Best Local Similarity: 98.02% Mismatches: 2  
 Query Match: 46.41% Indels: 0  
 DB: 2 Gaps: 0

US-09-595-947c-10 (1-214) x AC011010 (1-170896)

QY 114 ProAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrp 133  
 DB 30631 CCAGACGACGGAAGCTCACCAGATCGAGAGCTGGCTCGCCCACTACATCTGG 30572  
 QY 134 AlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAlaLeuGluProProAla 153  
 DB 30571 GCGTGACTCAACGCTGCGCATAGCGACACACAGCTTGACGCGCTGGAGCGCGCGG 30512  
 QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173  
 DB 30511 CCGCAGCTCGGGGAGCTGGGCGAGCCCGGCTCCCGGGGACTGGGGTCCCTCTAC 30452  
 QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluArgProGly 193  
 DB 30451 TCCCGAGTCTCCAGGCTGGCAGCTGAGTCCCGCGCTGCTGGAGGAGCGACCGGG 30392  
 QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySerLeuAlaPheSerAspPhe 213  
 DB 30391 CTGCTGGGGGCCACCTTTTCGCTGCTGTAGCCCGGAGCTGTGGCTTTCTCAGATTT 30332  
 QY 214 Leu 214  
 DB 30331 CTG 30329  
 RESULT 13

GGA012659 790 bp DNA linear VRT 03-JAN-2001  
 LOCUS Gallus gallus ngn2/ath4a gene.  
 DEFINITION AJ012659  
 ACCESSION AJ012659.1 GI:3892740  
 VERSION atonal protein; bHLH transcription factor; neurogenin.  
 KEYWORDS chicken.  
 SOURCE ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 790)  
 AUTHORS Matter-Sadzinski, L., Matter, J. M., Ong, M. T., Hernandez, J. and Ballivet, M.  
 TITLE Specification of neurotransmitter receptor identity in developing retina: the chick *ATH5* promoter integrates the positive and negative effects of several bHLH proteins  
 JOURNAL Development 128 (2), 217-231 (2001)  
 MEDLINE 21064448  
 PUBMED 11124117  
 REFERENCE 2 (bases 1 to 790)  
 AUTHORS Ballivet, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-1998) Ballivet M., Biochemistry, University of Geneva, 30, quai Ernest Ansermet, 1211 Geneva 4, SWITZERLAND  
 FEATURES  
 source  
 1..790  
 /organism="Gallus gallus"  
 /db\_xref="taxon:9031"  
 /clone="p56.1"  
 /clone.lib="lambda L47, homemade"  
 116..757  
 /gene="ngn2"  
 116..757  
 /note="ath4a"  
 /gene="ngn2"  
 /function="bHLH transcription factor"  
 /note="neurogenin 2"  
 /codon\_start=1  
 /evidence=experimental  
 /product="atonal-like protein 4a"  
 /protein\_id="CAAL0105.1"  
 /db\_xref="GI:3892741"  
 /db\_xref="SPTREMBL:O9YH82"  
 /translation="MPVKASPPAAEDLELLRLASPPASASLPSSAGBDEDEEDG  
 RPRRLQEGARRAGQPPRAATAETAKRIKRRUKANNRRNRNHNNAALDLR  
 DVLPTPEDAKLTKIETLRFYHNYIALTETLRAGARLGGADAAAPGAAAGSPSP  
 ASSWSGGASPPASPVACTLSPGSPAGSADAEHWPGRFAPPPPHRL"  
 BASE COUNT 91 a 351 c 283 g 65 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.28e-11 Length: 790  
 Score: 388.50 Matches: 104  
 Percent Similarity: 54.31% Conservations: 22  
 Best Local Similarity: 44.83% Mismatches: 63  
 Query Match: 34.47% Indels: 43  
 DB: 5 Gaps: 10  
 US-09-595-947c-10 (1-214) x GGA012659 (1-790)  
 QY 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22  
 DB 47 CCAGCGCCCCCGACCGCCCGACGCGCC---GTCAGCACGCGCGCGCCCTCTCTCCC 103  
 QY 23 Arg-----AlaSerGluAspGluVal 29  
 DB 104 CGCCCCCGCAGGATGCGGTGAAGCGGAGAGCGCGCGCGCGGAGGAGCACTG 163  
 QY 30 ThrCysProThrSerAlaProProSerProThrArgThr----ProGlyAsnCysAlaGlu 48  
 DB 164 CTGCTGTGCGCTCGCGCTCGCGCGCCCTCGCGCTCGCTGCGCTCGCGCGCGCGGAG 223





SOURCE  
ORGANISM  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS  
1 (bases 1 to 105572)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Anaratunga, H.C., Are, J.R., Avele, M., Banks, T.,  
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brivea, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Devila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwater, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Herrandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
Homs, J., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H.,  
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokukwu, S., Ogih, M., Okwuonu, G.,  
Oragunye, N., Oriado, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

TITLE  
JOURNAL  
Direct Submission  
Unpublished  
2 (bases 1 to 105572)  
Worley, K.C.

REFERENCE  
AUTHORS  
TITILE  
JOURNAL  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 105572)  
Worley, K.C.

REFERENCE  
AUTHORS  
TITILE  
JOURNAL  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:18701957.

COMMENT  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GPAB  
Center clone name: CH230-235K10  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329

Consensus quality: 42382 bases at least Q40  
Consensus quality: 45708 bases at least Q30  
Consensus quality: 49449 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/GenBank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/GenBank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 55 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1274: contig of 1274 bp in length  
\* 1375 1374: gap of unknown length  
\* 1375 2888: contig of 1514 bp in length  
\* 2889 2988: gap of unknown length  
\* 2989 4406: contig of 1418 bp in length  
\* 4407 4506: gap of unknown length  
\* 4507 6096: contig of 1590 bp in length  
\* 6097 6196: gap of unknown length  
\* 6197 7656: contig of 1460 bp in length  
\* 7657 7756: gap of unknown length  
\* 7757 9085: contig of 1329 bp in length  
\* 9086 9185: gap of unknown length  
\* 9186 10429: contig of 1244 bp in length  
\* 10430 10539: gap of unknown length  
\* 10540 11704: contig of 1175 bp in length  
\* 11705 11804: gap of unknown length  
\* 11805 13085: contig of 1281 bp in length  
\* 13086 13185: gap of unknown length  
\* 13186 14926: contig of 1741 bp in length  
\* 14927 15026: gap of unknown length  
\* 15027 16549: contig of 1523 bp in length  
\* 16550 16649: gap of unknown length  
\* 16650 18165: contig of 1516 bp in length  
\* 18166 18265: gap of unknown length  
\* 18266 19724: contig of 1459 bp in length  
\* 19725 19824: gap of unknown length  
\* 19825 20829: contig of 1005 bp in length  
\* 20830 20929: gap of unknown length  
\* 20930 22321: contig of 1392 bp in length  
\* 22322 22421: gap of unknown length  
\* 22422 23592: contig of 1171 bp in length  
\* 23593 23692: gap of unknown length  
\* 23693 25176: contig of 1484 bp in length  
\* 25177 25276: gap of unknown length  
\* 25277 27028: contig of 1752 bp in length  
\* 27029 27128: gap of unknown length  
\* 27129 28815: contig of 1687 bp in length  
\* 28816 28915: gap of unknown length  
\* 28916 30149: contig of 1234 bp in length  
\* 30150 30249: gap of unknown length  
\* 30250 31482: contig of 1233 bp in length  
\* 31483 31582: gap of unknown length  
\* 31583 33013: contig of 1431 bp in length  
\* 33014 33113: gap of unknown length  
\* 33114 34306: contig of 1193 bp in length  
\* 34307 34406: gap of unknown length  
\* 34407 35732: contig of 1326 bp in length  
\* 35733 35832: gap of unknown length  
\* 35833 36987: contig of 1155 bp in length  
\* 36988 37087: gap of unknown length  
\* 37088 38488: contig of 1401 bp in length  
\* 38489 38588: gap of unknown length  
\* 38589 40099: contig of 1511 bp in length  
\* 40100 40199: gap of unknown length  
\* 40200 41975: contig of 1776 bp in length  
\* 41976 42075: gap of unknown length  
\* 42076 43414: contig of 1339 bp in length  
\* 43415 43514: gap of unknown length  
\* 43515 45944: contig of 2430 bp in length



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 21:14:33 ; Search time 252 Seconds  
(without alignments)  
1912.411 Million cell updates/sec

Title: US-09-595-947c-10

Perfect score: 1127

Sequence: 1 MTPQSGAPTQVTRPERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn2.model -DEV=xlh  
-Q=/cgn2\_1/uspto\_spool/US09595947/runat\_07042003\_160404\_20055/app\_query.fasta\_1.391  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptn -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09595947.ecgn\_1\_1\_263 -runat\_07042003\_160404\_20055 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1105	98.0	5340	21	AAC61089	Human neurogenin 3
2	849	75.3	1491	19	AAV42512	CDNA encoding a no
3	819	72.7	804	19	AAV27050	Mouse neurogenin 3
4	819	72.7	804	21	AAZ51981	Murine neurogenin-3
5	819	72.7	861	22	AAF27266	Murine neurogenin 3
6	819	72.7	1861	21	AAC61090	Murine neurogenin
7	819	72.7	5567	22	AAF27254	Mouse atonal homol
8	519	46.1	592	24	ABQ49522	Oligonucleotide fo
9	519	46.1	592	24	ABQ49523	Oligonucleotide fo
10	467	41.4	592	24	ABQ49524	Oligonucleotide fo
11	467	41.4	592	24	ABQ49525	Oligonucleotide fo
12	388.5	34.5	790	22	AAF27264	Chicken atonal hom
13	385.5	34.2	1074	22	AAF27263	Chicken atonal hom
14	378.5	33.6	1268	18	AAT74891	Human neurogenic d
15	378.5	33.6	1268	19	AAV42932	DNA encoding human
16	375.5	33.3	1385	19	AAV27049	Mouse neurogenin 2
17	375.5	33.3	1385	21	AAZ51980	Murine neurogenin-3
18	375.5	33.3	1385	22	AAF27269	Mouse neurogenin 2
19	370.5	32.9	1412	22	AAF27255	Mouse atonal homol
20	370.5	32.9	1412	22	AAF27273	Mouse atonal homol
21	370	32.8	1527	19	AAV27045	Rat neurogenin 1 9
22	370	32.8	1527	21	AAZ51976	Rat neurogenin-1 (
23	365	32.4	1332	19	AAV42938	DNA encoding murin
24	365	32.4	1332	18	AAT74894	Mouse neurogenic d
25	359.5	31.9	738	19	AAV27046	Murine neurogenin 1
26	359.5	31.9	738	21	AAZ51977	Murine neurogenin-
27	322.5	28.6	1312	19	AAV27047	Xenopus neurogenin
28	322.5	28.6	1312	21	AAZ51978	Xenopus X-ngnr-1a
29	304	27.0	1277	19	AAV27048	Xenopus neurogenin
30	304	27.0	1277	21	AAZ51979	Xenopus X-ngnr-1b
31	297.5	26.4	778	24	ABQ16590	Oligonucleotide fo
32	297.5	26.4	778	24	ABQ16591	Oligonucleotide fo
33	248	22.0	778	24	ABQ16592	Oligonucleotide fo
34	248	22.0	778	24	ABQ16593	Oligonucleotide fo
35	241	21.4	2161	23	ABL13239	Drosophila melanog
36	241	21.4	4161	23	ABL13238	Drosophila melanog
37	238.5	21.2	1550	22	AAF27276	Mouse atonal homol
38	238.5	21.2	1957	22	AAF27256	Mouse atonal homol
39	234.5	20.8	1344	21	AA60796	Human MATH-2 encod
40	234.5	20.8	1344	22	AAH76705	Human atonal prote
41	233.5	20.7	1588	22	ABH82993	Human transcriptio
42	231.5	20.5	1550	21	AAA62681	Human NeuroD2 gene
43	230	20.4	1535	18	AAT74890	Human neurogenic d
44	230	20.4	1535	19	AAV42931	DNA encoding human
45	230	20.4	1951	18	AAT74893	Mouse neurogenic d

ALIGNMENTS

RESULT 1

AAC61089  
ID AAC61089 standard; DNA; 5340 BP.  
XX AAC61089;  
AC AAC61089;  
XX  
DT 05-FEB-2001 (first entry)  
XX Human neurogenin 3 (Ngn3) genomic DNA sequence.

Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;  
Islet cell precursor identification; diabetes mellitus; human; ds.

Homo sapiens.

Key Location/Qualifiers  
CDS 3022..3666  
FT CDS /\*tag= a

FT /product= "Ngn3"  
 XX /note= "Neurogenin 3"  
 PN W0200059936-A1.  
 XX 12-OCT-2000.  
 PD  
 XX  
 PF 28-MAR-2000; 2000WO-US08436.  
 XX  
 PR 06-APR-1999; 9905-0128180.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 PI German MS, Lin J;  
 XX  
 XX WPI; 2000-664989/64.  
 DR P-PSDB; AAY85617.  
 XX  
 XX Novel human neurogenin 3 polypeptides and polynucleotides encoding  
 PT them, useful for diagnosis, prevention and treatment of diabetes  
 PT mellitus and to identify individuals at risk of diabetes -  
 XX  
 XX Claim 6; Page 46-48; 54pp; English.  
 XX  
 CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3  
 CC protein AAY85617. The Ngn3 gene is located at chromosome position  
 CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and  
 CC protein sequences, and includes an antibody recognising the Ngn3 protein.  
 CC Also included in the invention is a method for identifying an islet cell  
 CC precursor, the method involves analysing a cell for the expression of the  
 CC Ngn3 gene product, where detection of the product is indicative of an  
 CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic  
 CC reagent for detecting (in a subject) a predisposition to a defect in  
 CC pancreatic islet cell function or formation associated with a defect in  
 CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell  
 CC precursor cells expressing Ngn3, and to alter cellular differentiation in  
 CC culture in vivo to produce new beta-cells to treat patients with diabetes  
 CC mellitus.  
 XX  
 SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.6e-59 Length: 5340  
 Score: 1105.00 Matches: 211  
 Percent Similarity: 99.07% Conservativeness: 1  
 Best Local Similarity: 98.60% Mismatches: 2  
 Query Match: 98.05% Indels: 0  
 DB: 21 Gaps: 0

US-09-595-947c-10 (1-214) x AAC61089 (1-5340)

QY 1 MethThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 |||||  
 DB 3022 ATACGCGCTCAAGCTCGGTGGCGCCACGTGCTCAAGTACCGCTGAGACGGAGCGGTCC 3081  
 |||||  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 |||||  
 DB 3082 TTCCCCAGACCTCGGAAGACGAAGTACCTGCCCGCCACGTGCGCCCGCCAGCCCGCCACT 3141  
 |||||  
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60  
 |||||  
 DB 3142 CGCACACGGGGGAACCTCGCAGAGGCGGAAGAGGAGGCTGCCGAGGGGCCCCGAGGAG 3201  
 |||||  
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 |||||  
 DB 3202 CTCGGGCACGGCGGGGACGCGCGGCTAAGACGAGTGGCAGTGGCAGCAGCAG 3261  
 |||||  
 QY 81 ArgArgSerArgArgLysLysAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 |||||  
 DB 3262 CGACGGAGTGGGAAAGAGGCCAACGCGGAGGCGCAATCGAATGACACACCTCAAC 3321  
 |||||  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120  
 |||||

DB 3322 TCGGCACTGACGCCCTGCGCGGTGCTCTGCCACCTTCCACAGACGAGAGCTCACC 3381  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 |||||  
 DB 3382 AAGATCGAGACGCTGCGCTTCGCCACCACTAGATCTGGCGCTGACTCAACGCTGCGC 3441  
 |||||  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 |||||  
 DB 3442 ATAGCGGACCACTGTTAGCGCTGGAGCGCGCGGCGGCTGCGGGAGCTGGGC 3501  
 |||||  
 QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180  
 |||||  
 DB 3502 AGCCAGCGGTTCCCGCGGACTGGGGTCTCTACTCCAGCTCCAGGCTGGC 3561  
 |||||  
 QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200  
 |||||  
 DB 3562 AGCTGAGTCCCGCGCTGCTGGAGAGCGACCGCGGCTGTGGGGGCGCACCTCTTC 3621  
 |||||  
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 |||||  
 DB 3622 GCCTGCTTGAGCCAGGACGCTGGCTTCTCAGATTTCTG 3663

RESULT 2  
 AAV42512  
 ID AAV42512 standard; cDNA; 1491 BP.  
 XX  
 AC AAV42512;  
 XX  
 DT 05-OCT-1998 (first entry)  
 XX  
 DE cDNA encoding a novel BHLH protein designated RELAX.  
 XX  
 KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;  
 KW control; gene expression; transcriptional activator; targeting;  
 KW protein expression; central nervous system; CNS; treatment;  
 KW nervous system disorder; ss.  
 XX  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 459..1103  
 FT /\*tag= a  
 FT /product= RELAX  
 XX  
 PN W09827206-A2.  
 XX  
 PD 25-JUN-1998.  
 XX  
 PF 19-DEC-1997; 97WO-FR02368.  
 XX  
 PR 19-DEC-1996; 96FR-0015651.  
 XX  
 PA (RHON ) RHONE-POULENC RORER SA.  
 XX  
 PI Mallet J, Ravassard P, Icard-Liepkalns C;  
 XX  
 DR WPI; 1998-362775/31.  
 DR P-PSDB; AAW62991.  
 XX  
 PT Basic helix-loop-helix polypeptide and related nucleic acid - with  
 PT transcriptional activity, for targeting expression of genes to  
 PT central nervous system and treatment of nervous disease  
 XX  
 XX Claim 6; Page 20; 28pp; French.  
 CC The present sequence encodes a basic helix-loop-helix (BHLH) type  
 CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.  
 CC The protein is used to control and participate in gene expression,  
 CC by acting as transcriptional activator, strictly dependent on the  
 CC presence of an intact E box (CANNTG), particularly for targeting  
 CC expression of proteins to the central nervous system (CNS). The  
 CC nucleic acid sequence can be used to treat nervous system disorders,  
 CC and antisense sequences can be used to control mRNA transcription.

117 *Ray* *STATISTICS/METHODS*

Pred. No.:	1.27e-42	Length:	804
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2
DB:	19	Gaps:	2

US-09-595-947C-10 (1-214) x AAV27050 (1-804)

Qy	1	MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer	20
Db	160	ATGGCGCTCATCCTTTGGATGGGCTCACCATCAAGTGCCCGAGAGACACAACT	219
Qy	21	PheProArgAlaSerGluAspGlyValThrCysProThrSerAlaProProSerProThr	40
Db	220	TTTCCCGGAGCTCGGACACGAGTGCTCAGTTCCAAATCCACCCCACTAGCCCACT	279
Qy	41	ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys	60
Db	280	CTCATACTAGGGACTCTCCGAGACAGAAAGTGGTGACTGCCGAGGACCTCAGGAAG	339
Qy	61	LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln	80
Db	340	CTCGCGCCCGAGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAACAG	399
Qy	81	ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn	100
Db	400	CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACTCAAC	459
Qy	101	SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr	120
Db	460	TCGGCGCTGATGGCTGCGCGGTGTCTCTCCCACTTCCTCCGATGACGCGCAAACTTACA	519
Qy	121	LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg	140
Db	520	AAGATCGAGACCTTCGCTTCGCCCAACACTAGCATCTGGCACTGACTCAGAGCTGGCG	579

QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
 Db 580 ATAGCGGACCACTTCATGGCCGAGCCCTGTGCC---TGTGGAGAGCTGGG 636  
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179  
 Db 637 AGCCCGGAGTGGCTCCAAACGGGAGCTGGGCTCTATCTACTCCCAAGG 696  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 Db 697 GGTAACTGAGCCCGCCGCTCATTTGGAGGAATTCCTCGCTGCGAGGTGCCAGCTCC 756  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 Db 757 CCATCTATCTGCTCCCGGAGCACTGGTGTCTCAGACTTCTTG 801

RESULT 4  
 AAZ51981  
 ID AAZ51981 standard; DNA; 804 BP.  
 AC AAZ51981;  
 XX  
 DT 04-JUL-2000 (first entry)  
 XX  
 DE Murine neurogenin-3 (NGN3) nucleic acid sequence.  
 XX  
 KW Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;  
 KW Phox2a protein; neuronal subtype-specific marker; growth factor;  
 KW neural differentiation; transplantation; neuronal dysfunction;  
 KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;  
 KW neuroprotective; nontropic; anticonvulsant; antiparkinsonian; vulnary;  
 KW cerebroprotective; immunosuppressant; antiinfectious; ds.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 160..804  
 FT /\*tag= a  
 FT /product= "Murine neurogenin-3 protein"  
 XX  
 PN WO200009676-A2.  
 XX  
 PD 24-FEB-2000.  
 XX  
 PF 13-AUG-1999; 99WO-US18525.  
 XX  
 PR 14-AUG-1998; 98US-0096630.  
 XX  
 PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 XX  
 PI Anderson DJ, Lo L;  
 XX  
 DR WPI: 2000-256250/22.  
 DR P-PSDB; AA70570.  
 XX  
 PT Inducing non-neuronal cells to differentiate into neurons and for  
 PT non-neuronal cells to express a neuronal subtype-specific marker,  
 PT comprising contacting the non-neuronal cells with a vector containing  
 PT neurogenin nucleic acid -  
 XX  
 PS Claim 1; Fig 1J; 76pp; English.  
 XX  
 CC The patent discloses a method for inducing non-neuronal cells (NNC) to  
 CC differentiate into neurons and for NNCs to express a neuronal subtype  
 CC -specific marker. Transformed host cells are used as sources of neuronal  
 CC and other growth factors; in culture for screening compounds that  
 CC modulate neural differentiation or as sources of recombinantly produced  
 CC neurogenins and Phox2a proteins for use in transplantation. The cells  
 CC also have a variety of in vivo uses, e.g. for transplantation at sites of  
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to  
 CC optical or auditory nerve damage, brain or spinal cord injuries, and  
 CC neurodegenerative disorders e.g. Alzheimer's disease. The present  
 CC sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.

CC NNCs differentiate into neurons through the recombinant expression of a  
 CC transcription factor that induces a core program of neurogenesis. Forced  
 CC expression of murine NGN3 can elicit expression of at least some neuronal  
 CC phenotypic markers even in NNCs.

XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

#### Alignment Scores:

Pred. No.: 1,27e-42 Length: 804  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.33% Conservative: 14  
 Best Local Similarity: 75.81% Mismatches: 36  
 Query Match: 72.67% Indels: 2  
 DB: 21 Gaps: 2

US-09-595-947C-10 (1-214) x AAZ51981 (1-804)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 Db 160 ATGGCGCTCATCCCTTGGATGGCTCACCATCCAACTGCCAGAGACACACACCT 219  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 Db 220 TTTCCCGGAGCTCGGACCAAGTCTCAGTTCCAAATCCACCCACCTAGCCCACT 279  
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60  
 Db 280 CTCACTACCTAGGACTGCTCCGAGCAAGAGTGGGTACTGCCGAGGAGCTCGAGGAG 339  
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 Db 340 CTCCGCGCGGACGCGGAGGCGCAACAGGCCCAAGAGCGAGTGGCACTCAGCAACAG 399  
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 Db 400 CGAAGAAGCGCGCGCAAGAAGGCAATGATCGGAGGCGCAATGCGATGCAACCTCAAC 459  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120  
 Db 460 TCGGCGCTGGATCGCTGCGGCTGCTCCGCCACCTCCCGGATGACGCGCAACTTACA 519  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 Db 520 AAGATGAGACCTGCGCTTCGCCCAACACTACATCTGGCACTGAGTCAAGCTGCGC 579  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 Db 580 ATAGCGGACCACTTCATGGCCGAGCCCTGTGCC---TGTGGAGAGCTGGG 636  
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179  
 Db 637 AGCCCGGAGTGGCTCCAAACGGGAGCTGGGCTCTATCTACTCCCAAGG 696  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 Db 697 GGTAACTGAGCCCGCCGCTCATTTGGAGGAATTCCTCGCTGCGAGGTGCCAGCTCC 756  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 Db 757 CCATCTATCTGCTCCCGGAGCACTGGTGTCTCAGACTTCTTG 801

#### RESULT 5

AAF27266

ID AAF27266 standard; cDNA; 861 BP.

XX AAF27266;

AC AAF27266;

XX 24-APR-2001 (first entry)

XX Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;  
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;

KW cellular proliferation; cerebellar granule neuron; gene therapy;  
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;  
 KW transgenic animal; ss.  
 XX  
 OS Mus musculus.  
 XX WO200073764-A2.  
 PN  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 01-JUN-2000; 2000WO-US15410.  
 XX  
 PR 01-JUN-1999; 99US-0137060.  
 XX 19-JAN-2000; 2000US-0176993.  
 PR  
 XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PA  
 XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
 PI WPI: 2001-032190/04.  
 DR P-PSDB; AAB60359.  
 DR  
 XX  
 PT Therapeutic use of atonal-associated nucleic acids or amino acids, or  
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
 PT osteoarthritis and abnormal cell proliferation.  
 XX  
 PS Disclosure; Page -: 142pp; English.  
 XX  
 CC The invention relates to the use of atonal-associated nucleic acid or  
 CC amino acid sequence, or any of its homologues or orthologues as  
 CC therapeutic agents for the treatment of deafness, partial hearing loss,  
 CC vestibular effects due to damage or loss of inner hair cells,  
 CC osteoarthritis and abnormal cell proliferation. The invention also  
 CC encompasses methods of screening for compounds which affect the  
 CC expression of an atonal-associated nucleic acid sequence in an animal,  
 CC and a transgenic animal in which an allele of a native atonal-associated  
 CC gene is replaced by a heterologous nucleic acid sequence, thus  
 CC inactivating the atonal-associated allele. The nucleic acids or proteins  
 CC may be used in a method of treating an animal for hearing impairment,  
 CC joint disease, balance disorders, abnormal cell proliferation, or other  
 CC disease related to loss of a functional atonal-associated nucleic acid or  
 CC protein. They may particularly be used to treat an animal with a  
 CC deficiency in cerebellar granule neurons or their precursors, and may  
 CC also be used in promoting mechanoreceptive cell growth and generating  
 CC hair cells. The present sequence represents an atonal-associated nucleic  
 CC acid sequence referred to in the invention.  
 CC Note: The present sequence is not shown in the specification, but  
 CC was obtained from GenBank.  
 XX  
 SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

## Alignment Scores:

Pred. No.: 1.36e-42 Length: 861  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.3% Conservative: 14  
 Best Local Similarity: 75.8% Mismatches: 36  
 Query Match: 72.6% Indels: 2  
 DB: 22 Gaps: 2

US-09-595-947C-10 (1-214) x AAF27266 (1-861)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrArgSer 20  
 Db 160 ATGGCGCCTCATCTGGTGGCTGCACCATCAAGTGTCCCGACAGACACAAACCT 219  
 QY 21 PheProArgAlaSerGluThrValThrCysProThrSerAlaProSerProThr 40  
 Db 220 TTTCGGGAGCTCGGACACGACCAAGTGTCAAGTTCACATTCACCCACCTAGCCCACT 279  
 QY 41 ArgThrProGlyAsnGlyAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60  
 Db 280 CTCATACCTAGGACTGCTCCGAGGAGAGAGTGGGTGACTGCCGAGGACCTCGAGGAG 339

QY 61 LeuArgAlaArgArgGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 Db 340 CTCGCCGCCGACGCGGAGGCGCAACAGCGCCAGAGCGTGTGGACTCAGCAACAG 399  
 QY 81 ArgArgSerArgGlyLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100  
 Db 400 CGAAGAGCGCGCGCAAGAAGCCCAATGATCGGAGCGCAATCGCATCACACCTCAAC 459  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 Db 460 TCGGGCGTGGATCGCTGCGCGTGTCTCCGCCACCTTCCCGGATGACGCCAACTTACA 519  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 Db 520 AAGATCGAGACCTCGGCTTCGCCCACTACATCTGGGCACTGACTCAGACGCTGGCG 579  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 Db 580 ATAGCGGACACACAGCTTCTATGGCGCGGAGCCCTGTGCC---TGTGGAGAGCTGGG 636  
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179  
 Db 637 AGCCCGGAGGTGGCTCCAAACGGGAGCTGGGCTCTATCTACTCCCACTCTCCCAAGCG 696  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuGlyAlaThrSer 199  
 Db 697 GGTAACTGAGCCCGACGCGCTCATTTGGAGGAATTCCTCCCTGGCTGAGGTGCCCACTCC 756  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 Db 757 CCATCTCTCTGCTCCCGGAGCAGCTGGTGTCTCAGACTTCTTG 801  
 RESULT 6  
 AAC61090  
 ID AAC61090 standard; DNA; 1861 BP.  
 XX  
 AC AAC61090;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Murine neurogenin 3 (Ngn3) genomic DNA sequence.  
 XX  
 KW Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;  
 KW islet cell precursor identification; mouse; ds.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1093..1737  
 FT /tag= a  
 FT /product= "Ngn3"  
 FT /note= "Neurogenin 3"  
 XX  
 PN WO200059936-A1.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-US08436.  
 XX  
 PR 06-APR-1999; 99US-0128180.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI German MS, Lin J;  
 XX  
 DR WPI: 2000-664989/64.  
 DR P-PSDB; AAY85618.  
 XX  
 PT Novel human neurogenin 3 polypeptides and polynucleotides encoding  
 PT them, useful for diagnosis, prevention and treatment of diabetes  
 PT mellitus and to identify individuals at risk of diabetes -  
 XX  
 PS Claim 18; Page 49-50; 54pp; English.



XX The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3  
 CC protein AAY85617. The Ngn3 gene is located at chromosome position  
 CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and  
 CC protein sequences, and includes an antibody recognising the Ngn3 protein.  
 CC Also included in the invention is a method for identifying an islet cell  
 CC precursor, the method involves analysing a cell for the expression of the  
 CC Ngn3 gene product, where detection of the product is indicative of an  
 CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic  
 CC reagent for detecting (in a subject) a predisposition to a defect in  
 CC pancreatic islet cell function or formation associated with a defect in  
 CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell  
 CC precursor cells expressing Ngn3, and to alter cellular differentiation in  
 CC culture in vivo to produce new beta-cells to treat patients with diabetes  
 CC mellitus. The present sequence represents the murine Ngn3 genomic DNA  
 CC sequence.

XX SQ Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;

#### Alignment Scores:

Pred. No.:	3e-42	Length:	1861
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2
DB:	21	Gaps:	2

US-09-595-947C-10 (1-214) x AAC61090 (1-1861)

QY	1	MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer	20
Db	1093	ATGGCGCCTCATCCCTGGATGCTGCCTCACCATCCAGTGTCCCCAGACACACACCT	1152
QY	21	PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr	40
Db	1153	TTTCCCGGACCTCGGACACGAGTGTCTAGTTCCAAATCCACCACCTAGCCCACT	1212
QY	41	ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys	60
Db	1213	CTCATACCTAGGAGCTCTCCGAGACAGAAAGTGGTGACTCCGAGGAGCTCGAGGAG	1272
QY	61	LeuArgAlaArgArgGlyAlaArgSerArgProLysSerGluLeuAlaLeuSerLysGln	80
Db	1273	CTCGCGCCCGACGCGGAGGCGGCAACAGCCGCAAGAGCGAGTGGCACTCAGCAACAG	1332
QY	81	ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn	100
Db	1333	CGAAGAACCCGCGGCAAGAGCCCAATGATCGGAGCGCAATCGCATCGACACCTCAAC	1392
QY	101	SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr	120
Db	1393	TCGCGCTGTGATGCTGCTGCGGTGCTCTCCACCTTCCTCGGATGACGCAAACTTACA	1452
QY	121	LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg	140
Db	1453	AGATCGAGACCTGCTGCTGCGCCCACTACATCTGGGCACCTGATCAGAGCTCGCG	1512
QY	141	IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly	160
Db	1513	ATAGCGGACACACCTTCTATGCGCGAGCCCTCTGCGCC---TCTGGAGAGCTGGG	1569
QY	161	SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla	179
Db	1570	AGCCCCGAGGTGGCTCCAAACGGGAGCTGGGGCTCTATCTACTCCCGCAGTCTCCCAAGCG	1629
QY	180	GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer	199
Db	1630	GGTAACCTGAGCCCGCCGCTCATTTGGAGAAATTCCTGGCTGAGGTCGCCAGCTCC	1689
QY	200	SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu	214
Db	1690	CCATCTATCTGCTCCCGGAGCAGCTGGTGTCTTCAGACTTCTTG	1734

#### RESULT 7

AAAF27254	AAAF27254 standard; cDNA; 5567 BP.
ID	AAF27254
XX	AAF27254;
XX	24-APR-2001 (first entry)
DT	Mouse atonal homologue 5 (ATO5, Math4B) cDNA, SEQ ID NO:4.
XX	Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW	hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW	cellular proliferation; cerebellar granule neuron; gene therapy;
KW	mechanoreceptive cell growth; auditory; osteopathic; cytoskeletal;
KW	transgenic animal; ss.
XX	Mus musculus.
XX	WO200073764-A2.
XX	07-DEC-2000.
XX	01-JUN-2000; 2000WO-US15410.
XX	01-JUN-1999; 99US-0137060.
PR	19-JAN-2000; 2000US-0176993.
XX	(BAYU ) BAYLOR COLLEGE MEDICINE.
XX	Zoghbi HY, Beilen H, Birmingham N, Hassan B, Ben-Arie N;
XX	WPI; 2001-032190/04.
DR	P-PSDB; AAB60350.
XX	Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT	any of its homologs or orthologs, for the treatment of e.g. deafness,
PT	osteoarthritis and abnormal cell proliferation -
XX	Disclosure; Page -; 142pp; English.
XX	The invention relates to the use of atonal-associated nucleic acid or
CC	amino acid sequence, or any of its homologs or orthologues as
CC	therapeutic agents for the treatment of deafness, partial hearing loss,
CC	vestibular effects due to damage or loss of inner hair cells,
CC	osteoarthritis and abnormal cell proliferation. The invention also
CC	encompasses methods of screening for compounds which affect the
CC	expression of an atonal-associated nucleic acid sequence in an animal,
CC	and a transgenic animal in which an allele of a native atonal-associated
CC	gene is replaced by a heterologous nucleic acid sequence, thus
CC	inactivating the atonal-associated allele. The nucleic acids or proteins
CC	may be used in a method of treating an animal for hearing impairment,
CC	joint disease, balance disorders, abnormal cell proliferation, or other
CC	disease related to loss of a functional atonal-associated nucleic acid or
CC	protein. They may particularly be used to treat an animal with a
CC	deficiency in cerebellar granule neurons or their precursors, and may
CC	also be used in promoting mechanoreceptive cell growth and generating
CC	hair cells. The present sequence represents an atonal-associated nucleic
CC	acid sequence referred to in the invention.
CC	Note: The present sequence is not shown in the specification, but
CC	was obtained from GenBank.
XX	
SQ	Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;

#### Alignment Scores:

Pred. No.:	9.25e-42	Length:	5567
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2
DB:	22	Gaps:	2

US-09-595-947C-10 (1-214) x AAF27254 (1-5567)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 4923 ATGGCGCCTCATCCTTGATCGCTCACCATCCAGTGTCCCGAGACACAAACCT 4982  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 4983 TTTCGGGAGCCTCGGACACCAAGTGTCTCAGTTCATTCACCCCACTAGCCCACT 5042  
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60  
 DB 5043 CTCATCTAGGACTGCTCCGAAGCAGAGTGGGTGACTCGCGAGGACCTCGAGGAAG 5102  
 QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 5103 CTCGGCGCCGACGGGAGCGCACAGCGCCACAGAGCGAGTGGCACTCAGCAACAG 5162  
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100  
 DB 5163 CGAAGAAGCGCGGCAAGAAGCCATGATCGGGAGCGCAATCGCATGCACCACTCAAC 5222  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120  
 DB 5223 TCGGGCGTGGATCGCTGCGCGGTGCTTCCGCCACCTCCCGATGACGCCAACCTACA 5282  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 DB 5283 AGATCGAGACCTCGGCTTCGCCACACACTACATCTGGCACTGACTCAGACGCTGCGC 5342  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 DB 5343 ATAGCGGACACAGCTTCTATGGCCCGGAGCCGCCCTGTGCC---TGTGGAGAGCTGGG 5399  
 QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179  
 DB 5400 AGCCCGGAGGTGGCTCCCAACGGGAGCTGGGCTCTATCTACTCCCACTCTCCCAAGCG 5459  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 DB 5460 GGTAACTGAGCCCGACGCTCATTTGGAGGAATTCCTGGCTGCGAGTGGCCCACTCC 5519  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 DB 5520 CCATCTATCTGCTCCGGGAGCACTGGTGTCTCAGACTTCTTG 5564  
 RESULT 8  
 ID ABQ49522/c  
 AC ABQ49522 standard; DNA; 592 BP.  
 AC ABQ49522;  
 DT 12-JUL-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 OS Homo sapiens.  
 OS WO200218632-A2.  
 PN 07-MAR-2002.  
 PD 01-SEP-2001; 2001WO-EP10074.  
 PF 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX (EPIC-) EPICENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.  
 XX Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3.8e-24 Length: 592  
 Score: 519.00 Matches: 98  
 Percent Similarity: 83.67% Conservative: 25  
 Best Local Similarity: 66.67% Mismatches: 24  
 Query Match: 46.05% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-595-947C-10 (1-214) x ABQ49522 (1-592)  
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 443 ATAAACGCTCAACCTCGAATACGCTATCCAAATAACCCGTAACAAAGAAAGATCC 384  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 383 TTCCCAAAACCTCGAAACGAAATAACCTACCCACGCTCCGCCGCCCAACCCCACT 324  
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60  
 DB 323 CGCACACGAAACAACTACGCAAAACGAAACAAACAAACAAACAAACAAACAAACAA 264  
 QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 263 CTCGGAACAGCGCGCAACAAACGCAACCGACTTAAACACGAATTAACACATAACAA 204  
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeuAsn 100  
 DB 203 CGACGAAATCGACGAAACAAACCAACGACGCGGACGCAATCGAATACACACCTCAAC 144  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120  
 DB 143 TCGACACTAAACGCTCAGCGATATCTTACCACCTTCCCAACGACGCGAATCACTACC 84  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 DB 83 AAATCGAAACGCTAGCTTCGCCCACTACATCTAATCGCTAACTCAACGCTACGC 24  
 QY 141 IleAlaAspHisSerLeuTyr 147  
 DB 23 ATAACGAACCAACTTATAC 3

```

RESULT 9
ABQ49523
ID ABQ49523 standard; DNA; 592 BP.
XX
AC ABQ49523;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
This invention describes a novel method for determining the degree of
methylation of a particular cytosine in a motif 5'-CpG-3', present in a
genomic sample of DNA. The sample is treated chemically to convert
cytosine (C) but not methylated C, to uracil, then part of the genomic
DNA that contains the target C is amplified to form a labeled amplicon.
The amplicon is hybridised to two classes, each with at least one
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
and the degree of hybridisation to both classes is determined from the
label on the amplicon. From the ratio of labels hybridised to the two
classes of oligomers, the degree of methylation is calculated. The method
is used: (i) for diagnosis and/or prognosis of side effects of
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
of the central nervous, cardiovascular, gastrointestinal and respiratory
systems etc., particularly by detecting mutations or single nucleotide
polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
types and for investigating cell differentiation. The method allows the
methylation status of many C residues to be determined simultaneously.
ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
method for determining the degree of cytosine methylation described in
the disclosure of the invention.
XX
SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

Alignment Scores:
Pred. No.: 3.8e-24 Length: 592
Score: 519.00 Matches: 98
Percent Similarity: 83.67% Conservative: 25
Best Local Similarity: 66.67% Mismatches: 24
Query Match: 46.05% Indels: 0
DB: 24 Gaps: 0

US-09-595-947c-10 (1-214) x ABQ49523 (1-592)
Oy 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20

```

```

:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 150 ATAAGCGCTCAACCTCGAATACGCCCACTATCAAAATAACCGTAAACGAAGATCC 209
Oy 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 210 TTCCCAAAACCTCGAAAACGAATAATACCTACCCACGTCGCCGCCGCCCAACCCACT 269
Oy 41 ArgThrProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 270 CGCACAGAAAACCTACGCCAAAACGAAAAAAAATACTACCGAAAAACCCCGAAAAA 329
Oy 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 330 CTCCGAACACGACGCGAAAAAAGCAACCGACTTAAACAAAGTAATTAACACTAAACAACAA 389
Oy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 390 CGACGAATCGACGAAAAAACCACCAACGCGCAACCAATCGAATACACAACCTCAAC 449
Oy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 450 TCGACACTAAACGCCCTACGGGATATCTACCCACCTTCCCAAGGAGCGGAACTCACC 509
Oy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 510 AAAATCGAAACGCTACGCTTCGCCCAACACTACATCTAAACGCTAACTCAACGCTAGCG 569
Oy 141 IleAlaAspHisSerLeuTyr 147
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 570 ATAACGAACCACTTATAC 590

RESULT 10
ABQ49524
ID ABQ49524 standard; DNA; 592 BP.
XX
AC ABQ49524;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
This invention describes a novel method for determining the degree of
methylation of a particular cytosine in a motif 5'-CpG-3', present in a
genomic sample of DNA. The sample is treated chemically to convert
cytosine (C) but not methylated C, to uracil, then part of the genomic
DNA that contains the target C is amplified to form a labeled amplicon.
The amplicon is hybridised to two classes, each with at least one
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
and the degree of hybridisation to both classes is determined from the
label on the amplicon. From the ratio of labels hybridised to the two
classes of oligomers, the degree of methylation is calculated. The method
is used: (i) for diagnosis and/or prognosis of side effects of
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
of the central nervous, cardiovascular, gastrointestinal and respiratory
systems etc., particularly by detecting mutations or single nucleotide
polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
types and for investigating cell differentiation. The method allows the
methylation status of many C residues to be determined simultaneously.
ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
method for determining the degree of cytosine methylation described in
the disclosure of the invention.
XX
SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

```

CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB04121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;

Alignment Scores:  
 Pred. No.: 6.4e-21 Length: 592  
 Score: 467.00 Matches: 101  
 Percent Similarity: 73.94% Conservative: 4  
 Best Local Similarity: 71.13% Mismatches: 37  
 Query Match: 41.44% Indels: 0  
 DB: 24 Gaps: 0

US-09-595-947C-10 (1-214) x ABQ49524 (1-592)

Qy 6 SerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPheProArgAlaSer 25  
 Db 165 TCGGGTGGTATTGTTTAAAGTATTCGTGAGACGGCGGTTTTTTTAGAGTTTCG 224  
 Qy 26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45  
 Db 225 GAAGACGAAGTATGTTTACGTTTCGTTTACGTTTATTCGTATCGTACGGGGGAAT 284  
 Qy 46 CysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLysLeuArgAlaArg 65  
 Db 285 TCGGTAGAGCGGAGAGAGAGGATGTCGAGGGGTTTCGAGGAAGTTTCGGTACGGCC 344  
 Qy 66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArg 85  
 Db 345 GGGGACGTAGTCGGTTTAAAGACGAGTTGTTAGTAACTAGCAGCGAGTCGGCA 404  
 Qy 86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105  
 Db 405 AAGAAGGTTAAGCATCGGAGCGTAATCGAATGATATTTTAAATTCGGTATTGACGTT 464  
 Qy 106 LeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeu 125  
 Db 465 TTGCGCGGTTTGTGTTTATTTTATAGACGCGGAGTTTAAAGATCGAGACGTG 524  
 Qy 126 ArgPheAlaHisAsnTyrlleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisser 145  
 Db 525 CGTTTCGTTTAAATATATATTGGCGTTGATTAAACGTTTCGATACGCGATTATAGT 584  
 Qy 146 LeuTyr 147  
 Db 585 TTGTAC 590

RESULT 11

ID ABQ49525 standard; DNA; 592 BP.  
 XX AC ABQ49525;  
 XX DT 12-JUL-2002 (first entry)  
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB04121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

Alignment Scores:

Pred. No.: 6.4e-21 Length: 592  
 Score: 467.00 Matches: 101  
 Percent Similarity: 73.94% Conservative: 4  
 Best Local Similarity: 71.13% Mismatches: 37  
 Query Match: 41.44% Indels: 0  
 DB: 24 Gaps: 0

US-09-595-947C-10 (1-214) x ABQ49525 (1-592)

Qy 6 SerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPheProArgAlaSer 25  
 Db 428 TCGGGTGGTATTGTTTAAAGTATTCGTGAGACGGCGGTTTTTTTAGAGTTTCG 369  
 Qy 26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45  
 Db 368 GAAGACGAAGTATGTTTACGTTTCGTTTACGTTTATTCGTATCGGGGGAAT 309  
 Qy 46 CysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLysLeuArgAlaArg 65  
 Db 308 TCGGTAGAGCGGAGAGAGGATGTCGAGGGGTTTCGAGGAAGTTTCGGTACGGCC 249  
 Qy 66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArg 85  
 Db 66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArg 85



XX DE Chicken atonal homologue ngn1/ath4c cDNA, SEQ ID NO:18.

XX KW Atonal; homologue; orthologue; atonal-associated protein; deafness;

KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;

KW cellular proliferation; cerebellar granule neuron; gene therapy;

KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;

KW transgenic animal; ss.

XX OS Gallus gallus.

XX PN WO2000073764-A2.

XX PD 07-DEC-2000.

XX PF 01-JUN-2000; 2000WO-US15410.

XX PR 01-JUN-1999; 99US-0137060.

XX PR 19-JAN-2000; 2000US-0176993.

XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.

XX PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;

XX DR WPI: 2001-032190/04.

XX DR P-PSDB; AAB60356.

XX PT Therapeutic use of atonal-associated nucleic acids or amino acids, or

PT any of its homologs or orthologs, for the treatment of e.g. deafness,

PT osteoarthritis and abnormal cell proliferation.

XX PS Disclosure; Page : 142pp; English.

XX CC The invention relates to the use of atonal-associated nucleic acid or

CC amino acid sequence, or any of its homologues or orthologues as

CC therapeutic agents for the treatment of deafness, partial hearing loss,

CC vestibular effects due to damage or loss of inner hair cells,

CC osteoarthritis and abnormal cell proliferation. The invention also

CC encompasses methods of screening for compounds which affect the

CC expression of an atonal-associated nucleic acid sequence in an animal,

CC and a transgenic animal in which an allele of a native atonal-associated

CC gene is replaced by a heterologous nucleic acid sequence, thus

CC inactivating the atonal-associated allele. The nucleic acids or proteins

CC may be used in a method of treating an animal for hearing impairment,

CC joint disease, balance disorders, abnormal cell proliferation, or other

CC disease related to loss of a functional atonal-associated nucleic acid or

CC protein. They may particularly be used to treat an animal with a

CC deficiency in cerebellar granule neurons or their precursors, and may

CC also be used in promoting mechanoreceptive cell growth and generating

CC hair cells. The present sequence represents an atonal-associated nucleic

CC acid sequence referred to in the invention.

CC Note: the present sequence is not shown in the specification, but

CC was obtained from GenBank.

XX SQ Sequence 1074 Bp; 148 A; 449 C; 272 G; 205 T; 0 other;

Alignment Scores:

Pred. No.:	1,34e-15	Length:	1074
Score:	385.50	Matches:	104
Percent Similarity:	51.29%	Conservative:	15
Best Local Similarity:	44.83%	Mismatches:	60
Query Match:	34.21%	Indels:	53
DB:	22	Gaps:	10

US-09-595-947c-10 (1-214) x AAF27263 (1-1074)

QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerPro--- 39

DB 228 TTTCCCTCCCTTTCCCTCCCTCC-----CCTCAGCACCTTCCCTCCCTCCCTCCCT 275

QY 40 -----ThrArgThrProGlyAsnCysAlaGluAla 49

DB 276 GCGGCCCACTGAGCGGCTCGGCTTCTCTCCGAGAGATGCTCCGCGGAGCGGCGGAGCAGC 335

QY 50 GluGluGlyGlyCysArg-----GlyAlaProArgLysLeuArgAlaArgArgGly 66

DB 336 -----GGCGGGCTTTGGAGCGCGCGGAGCTCCCGGGAG---CGGCGGAGGAGACGC 386

QY 67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLys 86

DB 387 GCGCGTGGCGGGCGGACCGAGGCTTTGCTGCACACCTCAACAGGAGCGCGCGGTG 446

QY 87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerLalaLeuAspAlaLeu 106

DB 447 AAAGCCAGCAGCGGAGCGGAACCGCATGCACCTCAACCGCGCTGGATGAGCTC 506

QY 107 ArgGlyValLeuProThrPheProAspAlaLeuLysLeuThrLysIleGluThrLeuArg 126

DB 507 CGCAGCGTCTCGCGACCTTCCCGGACGACCACTCACCAGGAGCTTGGCGGAGCAGTGCCTC 566

QY 127 PheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146

DB 567 TTCGCTTACAACTACATCTGGGCCCTCTCCGAGAGCCTTCGTTTGGCGGAGCAGTGCCTC 626

QY 147 -----TyrAlaLeuGluProAlaProHisCysGlyGluLeuGlySerPro 162

DB 627 CCTCTCTCCCGCGCTTCCGCGGCGCGCGCGCGCC-----CCAGCGCC 671

QY 163 GlyGlyProProGlyAspTrpGlySerLeuTyrSerPro-----Val 176

DB 672 GCGAGGAGCGCGGTTCGTGGTGTCCAGGGTTCCCGCGCGCGCGCGCTCGCTCGCGGC 731

QY 177 SerGlnAlaGlySerLeuSerProAlaAlaSer----- 187

DB 732 TCGCGCTCGCGCGCGGAGCGCGCGCGCGCGCGCTCGGAGCTGCGGTACGTCCCTCGGAC 791

QY 188 ---LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla----- 201

DB 792 GCGCTCGCGCGCTTCCGCGGCTGCGCGCGCGCGCGCGCGCGCGCTCCCTCGCGCTAGCC 851

QY 202 -----CysLeuSerProGlySerLeuAlaPheSer 211

DB 852 TGCGCGTGGGTGCTCGTCCGCCCGCGCGCGCGCGCGCGCGCGCGCTCTCC 887

RESULT 14

AA74891

ID AA74891 standard; DNA; 1268 BP.

XX AC AA74891;

XX DT 02-OCT-1997 (first entry)

XX DE Human neurogenic differentiation protein (NeuroD3) DNA clone 20A1.

KW Neurogenic differentiation protein; NeuroD3 gene;

KW transcriptional activator; neuron; pancreas; gastrointestinal;

KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;

XX KW ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX CDS 55..768

XX FT /\*tag= a

XX XX

XX PN WO9716548-A1.

XX PD 09-MAY-1997.

XX PF 30-OCT-1996; 96WO-US17532.

XX PR 02-NOV-1995; 95US-0552142.

XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX PA (WEIN/) WEINTRAUB N.



Score:	378.50	Matches:	103
Percent Similarity:	51.0%	Conservative:	18
Best Local Similarity:	43.4%	Mismatches:	63
Query Match:	33.5%	Indels:	54
DB:	19	Gaps:	8

US-09-595-947C-10 (1-214) x AAV42932 (1-1268)

QY	4	GlnProSer-----GlyAlaProThrValGlnValThrArgGluThrGluArgSer	20
		:::	:::
Db	59	CAGCGCGCTTGACACTGCATCTCCGACCTCGACTGCCAGCAGCAGCAGCGCACTGACC	118
QY	21	PheProArgAlaSerGluAspGluValThrCys-----	-ProThr 33
		::	
Db	119	TATCCGGCTTCCTCACC -GACGAGAAGAACTGTGCCAGACTCCAACAGCAGCACCTCCGCT	177
QY	34	SerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGlyGly	53
		:::	
Db	178	TCGGGCGCGCGCGCGCGCGC-----	--- 201
QY	54	CysArgGlyAlaProArgLysLeuArgAla-----	--- 63
Db	202	--AGGGCGCGCCCATATCTCCCGGGGCTGTAGGTTCCAGGGSCAGCAGCACGACG	258
QY	64	-----ArgargGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys	79
		::    ::	
Db	259	CAGGAGGCGGGCGCGCGCGCGCGCGGTCGTCGAGCGGCTCGGAGCGGCTGCTGCATCG	318
QY	80	GlnArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMethHisAspLeu	99
		::    ::	
Db	319	CTGGCAGAGCGCGCGCTCAGGCCAAGCATCGGAGCGCAACGCATGCACAACATTG	378
QY	100	AsnSerAlaLeuaspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeu	119
		::    ::	
Db	379	AACGCGGCCCTGGACGCACATGCGCAGCGTCTGCCCTCGTTCCCCCAGCAGCACCAAGCTC	438
QY	120	ThrLysIleGluThrLeuArgPheAlaHisAsnTyrlTrpAlaLeuThrGlnThrLeu	139
		::    ::	
Db	439	ACCAAATCAGAGCGCTGCGCTTCGCTACAACTACATCTGGGCTCTGCCGAGACACHTG	498
QY	140	ArgIleAlaSpHisSerLeu-----	--- TyrAlaLeuGluPropro 152
Db	499	CGCCTGGCGGATCAAGGCTGCCGGAGGGGTCGCCGGAGCGGCTCTCGCGCGCGCAG	558
QY	153	AlaProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAsp-----	-Trp 169
Db	559	TGCGTCCCGCTGC-----CTGCGCGGTCCCGCCAGCGCCGCCAGCAGCGGAGTCTCGG	612
QY	170	GlySer-----LeuTy rSerProValSerGlnAlaGlySerLeuSerProAlaAla	186
		::    ::	
Db	613	GGCTCAGGTGCGCGCGCGCTCCCGCTCTCTGACCCCAGTAGCCAGCGCGCTCCGAA	672
QY	187	SerLeuGluGluArgProGly-LeuLeuGlyAlaThrSerSerAlaCys	202
		::    ::	
Db	673	GACTTCACCTACGCCCGCGCGACCTGTTTCTCTCTCCCAAGCTCGC	721

Search completed: April 8, 2003, 23:23:13  
Job time : 260 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: April 8, 2003, 23:15:09 ; Search time 564 Seconds  
(without alignments)  
332.826 Million cell updates/sec

Title: US-09-595-947C-10  
Perfect score: 1127  
Sequence: 1 MTPQSGAGTVQVTRERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 593429 seqs, 438583890 residues  
Total number of hits satisfying chosen parameters: 1186858  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09595947/runat\_07042003\_160406\_20125/app\_query.fasta\_1.391  
-DB=Published Applications\_NA -OFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LISH=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09595947 -ECGN\_1\_1\_93=runat\_07042003\_160406\_20125  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1105	98.0	5340	10	US-09-817-360-1
2	819	72.7	645	9	US-10-004-717-4
3	819	72.7	861	9	US-10-004-717-24
4	819	72.7	1861	10	US-09-817-360-3

5	388.5	34.5	790	9	US-10-004-717-20
6	385.5	34.2	1074	9	US-10-004-717-18
7	375.5	33.3	1385	9	US-10-004-717-30
8	370.5	32.9	1412	9	US-10-004-717-6
9	370.5	32.9	1412	9	US-10-004-717-37
10	370	32.8	1527	7	US-08-722-570-12
11	359.5	31.9	738	7	US-08-722-570-13
12	322.5	28.6	1312	7	US-08-722-570-14
13	304	27.0	1277	7	US-08-722-570-15
14	238.5	21.2	1550	9	US-10-004-717-43
15	238.5	21.2	1957	9	US-10-004-717-8
16	207.5	18.4	993	9	US-10-004-717-47
17	207.5	18.4	3261	9	US-10-004-717-12
18	207.5	18.4	3541	9	US-10-004-717-32
19	205.5	18.2	1056	9	US-10-004-717-10
20	205.5	18.2	1393	9	US-10-004-717-45
21	198.5	17.6	675	9	US-10-004-717-15
22	197.5	17.5	485	9	US-10-004-717-59
23	192.5	17.1	1065	9	US-10-004-717-1
24	192.5	17.1	1572	9	US-10-004-717-57
25	190	16.9	501	9	US-10-004-717-13
26	187	16.6	948	9	US-10-004-717-41
27	187	16.6	2993	9	US-10-125-237-56
28	183.5	16.3	907	9	US-10-004-717-65
29	183.5	16.3	17290	9	US-09-999-121-7
30	183.5	16.3	25760	9	US-09-999-121-13
31	177.5	15.7	938	9	US-10-004-717-39
32	164	14.6	485	10	US-09-833-381-66
33	163.5	14.5	450	9	US-10-004-717-22
34	162.5	14.4	446	10	US-09-833-381-505
35	161.5	14.3	1768	10	US-09-833-381-523
36	157	13.9	804	10	US-09-749-728B-62
37	157	13.9	249487	9	US-10-026-188-3
38	153	13.6	714	10	US-09-833-381-1316
39	153	13.6	1457	9	US-09-954-531-982
40	151	13.4	2466	12	US-10-04-090-251
41	149	13.2	1042	10	US-09-833-381-1305
42	149	13.2	2264	9	US-10-004-717-49
43	148	13.1	515	9	US-10-004-717-35
44	147	13.0	447	10	US-09-864-761-27737
45	144	12.8	1476	9	US-10-004-717-16

ALIGNMENTS

RESULT 1  
US-09-817-360-1  
; Sequence 1, Application US/09817360  
; Patent No. US20020015696A1  
; GENERAL INFORMATION:  
; APPLICANT: German, Michael S.  
; APPLICANT: Lin, Joseph  
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS  
; TITLE OF INVENTION: AND DELIVERY OF INSULIN  
; FILE REFERENCE: UCSF-129CIP  
; CURRENT APPLICATION NUMBER: US/09/817,360  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: 09/535,145  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: 60/128,180  
; PRIOR FILING DATE: 1999-04-06  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 5340  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-09-817-360-1  
Alignment Scores: 8.76e-97 Length: 5340  
Pred. No.: 1105.00 Matches: 211  
Score: 99.07%  
Percent Similarity: 99.07%  
Conservative: 1

Best Local Similarity: 98.60% Mismatches: 2  
 Query Match: 98.05% Indels: 0  
 DB: 10 Gaps: 0

US-09-595-947C-10 (1-214) x US-09-817-360-1 (1-5340)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 3022 ATGACGCTCAACCTCGGGTGGCCACTGCTCCAAAGTACCCGCTGAGAGCGAGCGTCC 3081
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
DB 3082 TTCCCCAGAGCTCGGAAGACGAGTACCTGCCCCACCTGCCGCCGCCAGCCCACT 3141
QY 41 ArgThrProGlyAsnGlyAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 3142 CGCACACGCGGGAACCTGCGCAGAGCGGAGAGGGAGGCTGCCGAGGGGCCCGGAGGAG 3201
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 3202 CTCGGGGACGCGCGGGGAGCGACCGCGGCTTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261
QY 81 ArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 3262 CGACGAGTCCGGAAAGAGCAACGACGCGGAGCGCAATCGAATGCACAACTCAAC 3321
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
DB 3322 TCGGCACTGGAGCGCCCTGGCGGTGTCTGCCACCTTCCAGACGACGCAAGCTCAC 3381
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 3382 AGATCGACAGCGTGGCGCTTCCGCCACAACTACATCTGGCGGTGACTCAAGCGTGGC 3441
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 3442 ATAGCGGACACAGCTGTACGCGCTGGAGCGCGCGGCGGCGGAGCTGGCG 3501
QY 161 SerProGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
DB 3502 AGCCAGGCGGTTCGCCGGGACTGGGGTCCCTCTACTCCCGACTCCCGAGGCTGGC 3561
QY 181 SerLeuSerProAlaAlaSerLeuGluArgGluArgProGlyLeuLeuGlyAlaThrSer 200
DB 3562 AGCTGAGTCCCGCCGCTGCTGGAGAGGAGCGGCGGCTGCTGGGGCCACCTCTTC 3621
QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 3622 GCCTGCTTCCAGCCAGGAGTCTGGCTTCTCAGATTCTTG 3663

```

## RESULT 2

```

US-10-004-717-4
; Sequence 4, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 645

```

TYPE: DNA  
 ORGANISM: Mus musculus  
 US-10-004-717-4

Alignment Scores: 2.34e-70 Length: 645  
 Pred. No.: 819.00 Matches: 163  
 Score: 819.00 Mismatches: 14  
 Percent Similarity: 82.33% Conservative: 36  
 Best Local Similarity: 75.81% Mismatches: 36  
 Query Match: 72.67% Indels: 2  
 DB: 9 Gaps: 2

US-09-595-947C-10 (1-214) x US-10-004-717-4 (1-645)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 1 ATGGCCCTCATCCCTTGGATGGCTCACCATCCAAAGTGTCCCGACAGACACAAACCT 60
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
DB 61 TTCCCGGAGCTCGGACCGACGAGTGTCTCAATTCACCCACCTAGCCCACT 120
QY 41 ArgThrProGlyAsnGlyAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 121 CTCATACCTAGGACTGTCTCCGAAGCAGAAAGTGGTGACTGCTCGGAGGAGCTCAGGAAG 180
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 181 CTCGCGCGCGGAGCGGCGCAACAGGCGCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 240
QY 81 ArgArgSerArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 241 CGAAGAGCGGCGGCGCAAGAGCGCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 300
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
DB 301 TCGGCGCTGGATGCGCTGGCGGTGTCTGCCACCTTCCCGGATGACGCGCAAACTTACA 360
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 361 AGATCGACAGCTTCTGCGCCACAACTACATCTGGGCACTGACTCAGAGCTCGCG 420
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 421 ATAGCGGACACAGCTTCTATGCGCGGAGCGGCGGCTGTGCGCC---TGTGGAGAGCTGGG 477
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
DB 478 AGCCCGGAGGTGGCTCCCAACGCGGAGTGGGGTCTATCTACTCTCTCTCTCTCTCTCTCT 537
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 538 GGTAACTGAGCCCGGCGCTCATTTGGAGGAATTCCTGCGCTGAGGTGCCAGCTCC 597
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 598 CCATCTATCTGCTCCCGGAGCAGCTGGTGTCTCAGACTTCTTG 642

```

## RESULT 3

```

US-10-004-717-24
; Sequence 24, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01

```

```
; PRIORITY APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: #0/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-24

Alignment Scores:
Pred. No.: 3,21e-70 Length: 861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 9 Gaps: 2

US-09-595-947C-10 (1-214) x US-10-004-717-24 (1-861)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 160 ATGGCGCTCATCCCTTGGATGCGCTCACCATCCAAAGTGTCCCGAGACACACACCT 219
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
Db 220 TTTCCTGGAGCCCTCGGACGACGAGAGTGTCTCAATTCACCTCCACCTAGCCCACT 279
QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyGlyGlyAlaProArgLys 60
Db 280 CTATACCTAGGACTGCTCCGAGACGAGAGTGGTGTCTCCGAGGAGCTCGAGGAG 339
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 340 CTCGCGCCGAGCGGAGGCGCACACAGGCGCAAGAGCGAGTGGCACTCAGCAACAG 399
QY 81 ArgArgSerArgGlyLysLysAlaAsnAspArgGluArgGlnArgMetHisAspLeuAsn 100
Db 400 CGAAGAGCGGCGCAAGAGGCGCAATGTCGGAGCGCAATCGCATGCAACCTCAAC 459
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 460 TCGCGCTGTGATGCTGCGCGGTGTCTCCACCTTCCTCCGATGACGCAACCTTACA 519
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 520 AAGATCAGACCTTGCCTTCGCCACAACTACATCTGGGCATCTGACCTGACGCTGCGC 579
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db 580 ATAGCGGACCACTTCTATGCGCGGAGCGCCCTGTGCC---TGTGGAGAGCTGGGG 636
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 637 AGCCCGGAGGTGGCTCCACGGGAGCTGGGGCTCTATCTACTCCCGAGTCTCCCAAGC 696
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 697 GGTAACCTGAGCCCGGAGGCTCATTTGGAGAAATTCCTGCGCTGCGCTGCGAGCTCC 756
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 757 CCATCCTATCTGCTCCGCGGAGCACTGGTGTCTCAGACTTCTTG 801

RESULT 4
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020013696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
```

```
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-817-360-3

Alignment Scores:
Pred. No.: 7,51e-70 Length: 1861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 10 Gaps: 2

US-09-595-947C-10 (1-214) x US-09-817-360-3 (1-1861)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 1093 ATGGCGCTCATCCCTTGGATGCGCTCACCATCCAAAGTGTCCCGAGACACACACCT 1152
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
Db 1153 TTTCCTGGAGCCCTCGGACGACGAGAGTGTCTCAATTCACCTCCACCTAGCCCACT 1212
QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyGlyGlyAlaProArgLys 60
Db 1213 CTCATACCTAGGACTGCTCCGAGACGAGAGTGGTGTCTCCGAGGAGCTCGAGGAG 1272
QY 61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 1273 CTCGCGCCGAGCGGAGGCGCACACAGGCGCAAGAGCGAGTGGCACTCAGCAACAG 1332
QY 81 ArgArgSerArgGlyLysLysAlaAsnAspArgGluArgGlnArgMetHisAspLeuAsn 100
Db 1333 CGAAGAGCGGCGCAAGAGGCGCAATGTCGGAGCGCAATCGCATGCAACCTCAAC 1392
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 1393 TCGCGCTGTGATGCTGCGCGGTGTCTCCACCTTCCTCCGATGACGCAACCTTACA 1452
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
Db 1453 AAGATCAGACCTTGCCTTCGCCACAACTACATCTGGGCATCTGACCTGACGCTGCGC 1512
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db 1513 ATAGCGGACCACTTCTATGCGCGGAGCGCCCTGTGCC---TGTGGAGAGCTGGGG 1569
QY 161 SerPro---GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
Db 1570 AGCCCGGAGGTGGCTCCACGGGAGCTGGGGCTCTATCTACTCCCGAGTCTCCCAAGC 1629
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 1630 GGTAACCTGAGCCCGGAGGCTCATTTGGAGAAATTCCTGCGCTGCGCTGCGAGCTCC 1689
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 1690 CCATCCTATCTGCTCCGCGGAGCACTGGTGTCTCAGACTTCTTG 1734

RESULT 5
US-10-004-717-20
```

Sequence 20, Application US/10004717  
Publication No. US20020192665A1  
GENERAL INFORMATION:  
APPLICANT: ZOGHEBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899054  
CURRENT APPLICATION NUMBER: US/10/004,717  
PRIOR FILING DATE: 2002-08-16  
PRIORITY FILING DATE: 2002-08-16  
PRIOR FILING DATE: 2000-06-01  
PRIOR FILING DATE: 2000-06-01  
PRIOR FILING DATE: 2000-01-19  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 20  
LENGTH: 790  
TYPE: DNA  
ORGANISM: chicken  
US-10-004-717-20

Alignment Scores:  
Pred. No.: 5 82e-29 Length: 790  
Score: 385.50 Matches: 104  
Percent Similarity: 54.31% Conservative: 22  
Best Local Similarity: 44.83% Mismatches: 63  
Query Match: 34.47% Indels: 43  
DB: Gaps: 10

US-09-595-947C-10 (1-214) x US-10-004-717-20 (1-790)

QY 3 ProGlnProSerGlyAlaProThrValGluThrArgGluThrGluArgSerPhePro 22  
Db 47 CCAGCGCCGCCCGACCGCCGGAGGC---GTGACACGCGCGCGCCCTCTCCCC 103  
QY 23 Arg-----AlaSerGluAspGluVal 29  
Db 104 CGCCCCCGCAGTATCGGTGAAGCGGAGAGCGCGCGCGCGCGGAGGACTG 163  
QY 30 ThrCysProThrSerAlaProSerProThrArgThr---ProGlyAsnCysAlaGlu 48  
Db 164 CTGCTGTCGCGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 223  
QY 49 AlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArg-----AlaArgArgGly 66  
Db 224 GAGGACGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 283  
QY 67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln-----Arg 81  
Db 284 GGCGCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATCAAG 343  
QY 82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101  
Db 344 CGGAGCGCGCGCGTGAAAGCACCAACCGGAGCGGAGCGGAGCGGAGCGGAGCGG 403  
QY 102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLys 121  
Db 404 CGCTGCGAGCGGCTGCGCGAGCTGCTGCCACCTTCCCAGGAGCGGAGCGGAGCG 463  
QY 122 IleGluThrLeuArgPheAlaHisAsnTyrlfetrpAlaLeuThrGlnThrLeuArgile 141  
Db 464 ATCGAGAGCGTGCCTTCGCCCAACAATACATCTGCGCGCTCACCGAGACGCTGCGGCT 523  
QY 142 AlaAspHlsSer----LeuTyraLeuGluProProAlaProHisCysGlyGluLeuGly 160  
Db 524 GCCGGGCGCGCGCTGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 583  
QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrserProValSerGlnAlagly 180

Db 584 AGCCCC---TCGCCCGCTCGTCTGG---AGCGCG 613  
QY 181 SerLeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSerSer 200  
Db 614 GCGCGCAGCGCGCGCGCTCC-----GCCTCGCCCTAC 646  
QY 201 AlaCys---LeuSerProGlySerLeuAlaPheSer 211  
Db 647 GCCTGCACCTTTATCGCCCGGCGAGCGCGCGCGCTCC 682

RESULT 6  
US-10-004-717-18  
Sequence 18, Application US/10004717  
Publication No. US20020192665A1  
GENERAL INFORMATION:  
APPLICANT: ZOGHEBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899054  
CURRENT APPLICATION NUMBER: US/10/004,717  
PRIOR FILING DATE: 2002-08-16  
PRIORITY FILING DATE: 2002-08-16  
PRIOR FILING DATE: 2000-06-01  
PRIOR FILING DATE: 2000-06-01  
PRIOR FILING DATE: 2000-01-19  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 18  
LENGTH: 1074  
TYPE: DNA  
ORGANISM: chicken  
US-10-004-717-18

Alignment Scores:  
Pred. No.: 1 58e-28 Length: 1074  
Score: 385.50 Matches: 104  
Percent Similarity: 51.29% Conservative: 15  
Best Local Similarity: 44.83% Mismatches: 60  
Query Match: 34.21% Indels: 53  
DB: Gaps: 10

US-09-595-947C-10 (1-214) x US-10-004-717-18 (1-1074)

QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerPro--- 39  
Db 228 TTTCCTCCCTTCCTCCCTCC-----CCTCAGCACCCTTCCCCCGCTCCCCCT 275  
QY 40 -----ThrArgThrProGlyAsnCysAlaGluAla 49  
Db 276 GCGCGCCACTGACGGCTCGCGCTCTCTCCCGAGGATCGCGGAGCGCGCGAGCAGC 335  
QY 50 GluGluGlyGlyCysArg-----GlyAlaProArgLysLeuArgAlaArgGly 66  
Db 336 -----GGCGCGGCTTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386  
QY 67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLys 86  
Db 387 GGCGGTGCGGCG 446  
QY 87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106  
Db 447 AAAGCAAACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 506  
QY 107 ArgGlyValLeuProThrPheProAspAlaLysLeuThrLysLysGluThrLeuArg 126  
Db 507 CGCAGCGCTCTGCGGACCTTCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 566  
QY 127 PheAlaHisAsnTyrlfetrpAlaLeuThrGlnThrLeuArgileAlaAspHlsSerLeu 146



```

: APPLICANT: YANG, QI
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
:
: TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
:
: TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
:
: FILE REFERENCE: P01999054
: CURRENT APPLICATION NUMBER: US/10/004,717
: CURRENT FILING DATE: 2002-08-16
: PRIOR APPLICATION NUMBER: 09/585,645
: PRIOR FILING DATE: 2000-06-01

```

```
OY 204 SerPro 205
Db 846 TCACCT 851

RESULT 10
US-08-722-570-12
; Sequence 12, Application US/08722570
; Publication No. US20030044887A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qiufu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1527 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-722-570-12

Alignment Scores:
Pred. No.: 7.16e-27 Length: 1527
Score: 370.00 Matches: 99
Percent Similarity: 52.97% Conservative: 17
Best Local Similarity: 45.21% Mismatches: 63
Query Match: 32.83% Indels: 40
DB: Gaps: 8

US-09-595-947c-10 (1-214) x US-08-722-570-12 (1-1527)
OY 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
Db 293 CCAGCAGCAACGCGGGAGCGACCTGTCAGTTTCCTCACCAGGAGGAGGACTGTGCCA 352
OY 23 ArgAlaSerGlu-----AspGluValThrCysProThr-----SerAlaPro 36
Db 353 GGCCTCCAGCCCTAGCTTCCACCTCAGGGCTCTCGTCCAGCCCGCAGGAGCGGCCCA 412
OY 37 ProSerPro-----ThrArgThr--ProGlyAsnCysAlaGluAlaGluGlyGlyCysA 55
Db 413 CCCTCTCCGGGCGATCGAAGCTCCCGTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 462
OY 55 .rgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyGlySerArgProLysSerGluL 75
Db 463 -----CGGGGGCGAGCGGAGGTCCGCGCGCGGTGGCGGTCCGAGG 502

75 euAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnA 95
Db 503 CGCTGCTGCACCTCGCTGCGGAGGAGCGCTCGCGTCAAGGCCAACGATCGCGAGCGCAACC 562
OY 95 rgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProA 115
Db 563 GTATGATAAACCTCAACGCTGCGCTGGACGCTCTCGCGAGCGTGTGCTGCTGCTGCTGCTG 622
OY 115 spAspAlaLysLeuThrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 135
Db 623 ACGACACCAAGCTCACCAGATTGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 582
OY 135 euThrGlnThrLeuArgLysLeuAlaAspHisSerLeu-----TyrA 148
Db 683 TGGCTGAGACACTGCGCTGCGAGATCAAGGGCTCCCGGGGGGGGGTGGCGGAGCGCC 742
OY 148 laLeuGluProProAlaProHisCysGlyGluLeuLysSerProGlyGlyProGlyA 168
Db 743 TCCTGCTCCCGCAGTGTGCTCCCTGC-----CTGCGCGGTCCCGGAGCGCGCGGT 796
OY 168 sp-----TrpGlySer-----LeuTyrSerProV 176
Db 797 ATACAGAGTCTGCGGGCTCGGGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856
OY 176 alSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluArgProGly 193
Db 857 TCTCTGACCCCGCAGTAGTCCCTCGGCTTCAGAAAGACTTACCTATGCGCGCGGT 909

RESULT 11
US-08-722-570-13
; Sequence 13, Application US/08722570
; Publication No. US20030044887A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qiufu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-722-570-13

Alignment Scores:
Pred. No.: 3.27e-26 Length: 738
Score: 359.50 Matches: 97
```





Qy	124	ThrLeuAArgPheAlaHisasnTyrIletrpAlaLeuThrGlnThrLeuArgIleAlaAsp	143
Db	562	ACCTTCGCTTTGCCACAACTACATCTGGGCTTATAGCAAACTTTGCGCTGGCCGAC	621
Qy	144	-----HisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGlu-----Leu	159
Db	622	CAGCTCGAGG-ATCTAC---TTCCACCCACACAGCAGCCATATTGGTACAGGACTCTCTA	677
Qy	160	GlySerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValserGlnAla	179
Db	678	TCCTTCCTGTA-GCCCTCTCTGGAGCTGCAGCTCGTCCCACTCTCCAACTCTTCGCACT	736
Qy	180	GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeu---LeuGlyAlaThr	198
Db	737	CCTTCCTCCCGACAGCCCTGCCAGCTCCACCTCGGACAGTATTGAGTACTGGCAGCCCT	796
Qy	199	SerSerAlaCysLeuSerProGlySerLeuAlaPhe	210
Db	797	CTGAGCTCGCTTGAAACCCTTCATGCTGCGCCTTT	832
RESULT 14			
US-10-004-717-43			
; Sequence 43, Application US/10004717			
; Publication No. US20020192665A1			
; GENERAL INFORMATION:			
; APPLICANT: ZOGHBI, HUDA Y.			
; APPLICANT: YANG, QI			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN			
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,			
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION			
; FILE REFERENCE: P01899054			
; CURRENT APPLICATION NUMBER: US/10/004,717			
; CURRENT FILING DATE: 2002-08-16			
; PRIOR APPLICATION NUMBER: 09/585,645			
; PRIOR FILING DATE: 2000-06-01			
; PRIOR APPLICATION NUMBER: 60/176,993			
; PRIOR FILING DATE: 2000-01-19			
; PRIOR APPLICATION NUMBER: 60/137,060			
; PRIOR FILING DATE: 1999-06-01			
; NUMBER OF SEQ ID NOS: 69			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 43			
; LENGTH: 1550			
; TYPE: DNA			
; ORGANISM: Mus musculus			
US-10-004-717-43			
Alignment Scores:			
Pred. No.:	3e-14	Length:	1550
Score:	238.50	Matches:	78
Percent Similarity:	39.29%	Conservative:	21
Best Local Similarity:	30.95%	Mismatches:	72
Query Match:	21.16%	Indels:	81
DB:	9	Gaps:	8
US-09-595-947C-10 (1-214) x US-10-004-717-43 (1-1550)			
Qy	19	ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro	36
Db	106	AGAAAGTTTGCTAGACAATGTGAGGACCAGAAACCAATTAAGAAACACAGAGAGCTTTCCA	165
Qy	37	-----ProSerProThrArgThrProGlyAsnCysAlaGluAla	49
Db	166	AAACAAGTTGCTCCTTCGAGGAAGAGCATTAAGAGGCCCTTCGGAGAGAAACCCAGAAA	225
Qy	50	Glu-----GlyGlyGlyCysArgGlyAlaProArg	59
Db	226	GAGAGGAGGAAGAAGACACAGAGGAGGAAGATGAGAAATGGC-----	267
Qy	60	LysLeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys	79
Db	268	---TTGTCGAAGAGGAGGGGCTCAGAAAACCAACCAACCAACCAACCTACGACTGGAAGG	324

QY 80 GlnArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99  
 Db 325 GTC---AAGTTCAGGAGAGAGAGCTAATGCGCGGAGAGGAAACCGGATGCACGGCCTC 381  
 QY 100 AsnSerAlaLeuaspAlaLeuArgGlyValLeuProThrPheProaspAlaLysLeu 119  
 Db 382 AATGATGCTCTGACAAATTTGCGAAAGTGGTCCCTGTTACTCTAAAACCCCAAAACTG 441  
 QY 120 ThrLysileGluThrLeuArgPheAlaHisAsnTyrileTrpAlaLeuThrGlnThrLeu 139  
 Db 442 TCCAAATAGAACTTACGACTGGCCAAATAATACATCTGGCAGCTTCTGAAATCTG 501  
 QY 140 ArgIle-----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 141  
 Db 502 AGGATTGGCAAGAGACCGGATCTGCTCACGTTGTCCTCAAAACTATGCAAAAGTCTTTCC 561  
 QY 141 ----- 141  
 Db 562 CAGCCAACTACAACTGTGGCAGGCTGCTTACAGCTCAACGCCAGAGATTTTCCTGATG 621  
 QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 153  
 Db 622 GGTCAAGGTGGGAGGCTGCCACACACAAAGTCAACCTACTCCACATTTCTACCCACCC 681  
 QY 154 ProHisCysGlyLeuLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173  
 Db 682 TACCACAGCCCTGAGCTGGCCACTCCGCCAGG-----CAT 717  
 QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluArgProGly 193  
 Db 718 GGGACTCTTGATATTCGAATTCGAATTCGAATTCGAATTCGAATTCGAATTCGAATTC 777  
 QY 194 LeuLeuGlyAlaThrSerAlaCysLeuSerPro 205  
 Db 778 TTCTATGAAGTACCTCCCTCCCTGAGTGTGCCAGCCCT 813

## RESULT 15

US-10-004-717-8  
 ; Sequence 8, Application US/10004717  
 ; Publication No. US20020192665A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZOGHBI, HUDA Y.  
 ; APPLICANT: YANG, QI  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
 ; FILE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
 ; FILE REFERENCE: P01899US4  
 ; CURRENT APPLICATION NUMBER: US/10/004,717  
 ; CURRENT FILING DATE: 2002-08-16  
 ; PRIOR APPLICATION NUMBER: 09/585,645  
 ; PRIOR FILING DATE: 2000-06-01  
 ; PRIOR APPLICATION NUMBER: 60/176,993  
 ; PRIOR FILING DATE: 2000-01-19  
 ; PRIOR APPLICATION NUMBER: 60/137,060  
 ; PRIOR FILING DATE: 1999-06-01  
 ; NUMBER OF SEQ ID NOS: 69  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 8  
 ; LENGTH: 1957  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 US-10-004-717-8

## Alignment Scores:

Pred. No.:	3,88e-14	Length:	1957
Score:	238.50	Matches:	78
Percent Similarity:	39.29%	Conservative:	21
Best Local Similarity:	30.95%	Mismatches:	72
Query Match:	21.16%	Indels:	81
DB:	9	Gaps:	8

QY 19 ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro 36  
 Db 182 AGAAAGTTTCTAGACAATGTGAGGACCAAGAAATAAAGAAACCCAGAGAGCTTTTCCA 241  
 QY 37 -----ProSerProThrArgThrProGlyAsnCysAlaGluAla 49  
 Db 242 AAACAAGTTTCTCTCTGAGGAAAGAGCAATTAAGAGGCCCTCGAGAGAAACCGAGAAA 301  
 QY 50 Glu-----GluGlyGlyCysArgGlyAlaProArg 59  
 Db 302 GAAGAGGAGGAAGAAGACAGAGAGGAAGATGAGATGGC----- 343  
 QY 60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79  
 Db 344 ---TTGTCCAGAAAGGAGGGGCTCAGGAAAAAAGACCAACCAACTACGACTGGAAGAAG 400  
 QY 80 GlnArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99  
 Db 401 GTC---AAGTTCAGGAGACAGGAAGCTAATGCGCGGAGAGGAACCGGATGCACGGCCTC 457  
 QY 100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProaspAlaLysLeu 119  
 Db 458 AATGATGCTCTGGCAATTTGCGAAAGTGGTCCCTGTTACTCTAAAACCCCAAAACTG 517  
 QY 120 ThrLysileGluThrLeuArgPheAlaHisAsnTyrileTrpAlaLeuThrGlnThrLeu 139  
 Db 518 TCCAAAAATAGAACTTTACGACTGGCCAAAAATACATCTGGCAGCTTCTGAAATTCG 577  
 QY 140 ArgIle----- 141  
 Db 578 AGGATTGGCAAGAGACCGGATCTGCTCACGTTGTCCTCAAAACTATGCAAAAGTCTTTCC 637  
 QY 141 ----- 141  
 Db 638 CAGCCAACTACAACTTGGTGGCAGGCTGCTTACAGCTCAACGCCAGAGTTTCTCTGATG 697  
 QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProAla 153  
 Db 698 GGTCAAGGTGGGAGGCTGCCACACACAAAGGTCAACCTACTCCACATTTCTACCCACCC 757  
 QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173  
 Db 758 TACCACAGCCCTGAGCTGGCCACTCCCCAGG-----CAT 793  
 QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluArgProGly 193  
 Db 794 GGCACCTCTTGATAATTCGAAGTCCATGAACCCCTACAACTACTGTCAGTGCATATGAATCC 853  
 QY 194 LeuLeuGlyAlaThrSerAlaCysLeuSerPro 205  
 Db 854 TTCTATGAAGTACCTCCCTCCCTGAGTGTGCCAGCCCT 889

Search completed: April 9, 2003, 09:58:37  
 Job time : 586 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 8, 2003, 22:56:59 ; Search time 1739 Seconds  
(without alignments)  
1993.004 Million cell updates/sec

Title: US-09-595-947C-10  
Perfect score: 1127  
Sequence: 1 MTQPSGAPVQVITRETERS.....LGATSSACLSFGSLAFSDFL 214

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xih  
-Q/cgn2.1/USPTO.spool/US09595947/runat\_07042003.160405.20073/app\_query.fasta.1.391  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=PCT -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09595947 -CGN\_1\_1396 -runat\_07042003.160405.20073 -NCPU=6 -ICPU=3  
-NO\_XLPAY -NO\_MMAB -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOP -DEV.TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	809	71.8	1540	11	AK008017 Mus muscu
2	676.5	60.0	593	17	AZ296526 RPCI-23-1
3	327	29.0	1098	14	BQ169355 STR0303
4	317.5	28.2	632	13	BQ101277 BJ010277
5	310.5	27.6	600	12	BG808248 2083-52 M
6	286.5	26.3	1039	12	BE780690 601469349
7	285.5	25.3	1037	17	CNS03VJ9 Tetraodon
8	280	24.8	629	13	BQ093114 BJ093114
9	278	24.7	595	13	BJ030202 BJ030202
10	268.5	23.8	804	9	AL540071 AL540071
11	267.5	23.7	401	10	AW147434 da02h12.y
12	251	22.3	588	10	AV673464 AV673464
13	249	22.1	490	10	AV995230 AV995230
14	243	21.6	814	14	BQ178789 UI-M-EVO-
15	239	21.2	1003	17	CNS02IB1 Tetraodon
16	236.5	20.1	742	14	BQ572268 UI-M-ED0-
17	226.5	20.1	770	14	BQ572426 UI-M-ED0-
18	224.5	19.9	688	12	BG699059 602678696
19	224.5	19.9	901	14	BQ424098 AGENCOURT
20	224.5	19.9	1967	11	BC022560 Homo sap1
21	222.5	19.7	953	9	AU067624 AU067624
22	219.5	19.5	724	12	BE783567 601471617
23	219	19.4	875	17	CNS02BRP Tetraodon
24	217.5	19.3	990	14	BQ086669 AGENCOURT
25	217.5	19.3	1065	12	BE777799 601463047
26	215	19.1	752	13	BI732893 603355021
27	215	19.1	809	12	BE783082 601470669
28	215	19.1	1009	12	BM559083 AGENCOURT
29	214	18.9	994	10	BB610830 BB610830
30	213.5	18.9	1055	9	AL535259 AL535259
31	212	18.8	600	12	BG802832 0191-69 M
32	212	18.8	651	12	BG296688 602393635
33	212	18.8	657	14	BM728742 UI-E-E01-
34	212	18.8	678	12	BG296579 602394490
35	212	18.8	692	13	BI737697 603358625
36	212	18.8	782	13	BI732663 603354913
37	212	18.8	835	12	BG298117 602396478
38	212	18.8	951	14	BQ891444 AGENCOURT
39	212	18.8	968	13	BI731880 603353625
40	212	18.8	1007	10	BB610207 BB610207
41	212	18.8	1017	14	BQ895731 AGENCOURT
42	212	18.8	1121	12	BG294629 602391973
43	212	18.8	2494	11	AK018781 Mus muscu
44	212	18.8	2495	11	AK005073 Mus muscu
45	209.5	18.6	879	13	BM450929 AGENCOURT

ALIGNMENTS

RESULT 1  
AK008017  
LOCUS  
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010001M19:atonal homolog 5 (Drosophila), full insert sequence.  
ACCESSION AK008017  
VERSION AK008017.1 GI:12841941  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to mRNA, clone:lib.RIKEN full-length enriched mouse cDNA library clone:2010001M19.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AK008017 1540 bp mRNA linear HTC 19-JAN-2002  
Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010001M19:atonal homolog 5 (Drosophila), full insert sequence.



Qy 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 |||||  
 Db 481 CGAAGACCGCGCGCAGAGAGCCAAATGATCGGAGCGCAATCGCATGCAACCTCAAC 540  
 |||||  
 Qy 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 |||||  
 Db 541 TCGCGCTGTGATGCGCTGCGCGGTGCTCTCCCACTTCGCGGATGACGCCAACTTACA 600  
 |||||  
 Qy 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140  
 |||||  
 Db 601 AAGTCGAGACCCCTGCGCTTCGCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 660  
 |||||  
 Qy 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
 |||||  
 Db 661 ATACGGACCAACACTGTATGCGCGGAGCCCTCTGTGCC---TGTGGAGACTGGGG 717  
 |||||  
 Qy 161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179  
 |||||  
 Db 718 AGCCCGGAGGTGCTCCACCGGGGACTGGGGCTCTATCTACTCCCACTCTCCAGCG 777  
 |||||  
 Qy 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 |||||  
 Db 778 GGTAACTGAGCCACCGGCTCTATGAGGAATTCCTGCGCTGCAAGTCCAGCTCC 837  
 |||||  
 Qy 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 |||||  
 Db 838 CCATCCTATCTGCTCCCGGGAGCACTGTGTCTCAGACTTCTTG 882  
 |||||

RESULT 2  
 AZ296526/c 593 bp DNA linear GSS 27-JUL-2000  
 LOCUS RPCI-23-160G18.TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18  
 DEFINITION , DNA sequence.  
 ACCESSION AZ296526  
 VERSION 1 GI:9538311  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 593)  
 Zhao.S., Nierman.W., Feldblum.T., Malek.J., Shatsman.S., Akinret  
 B., Levins.M., Mcgann.S., Tsegaye.G., Geer.K., Krol.M., de Jong,P.  
 and Fraser.C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 Other\_GSSs: RPCI-23-160G18.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
 or from Resea ch Genetics ([info@resgen.com](http://info@resgen.com)). BAC end page:  
[http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 160 row: 6 column: 18  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers  
 1. 593  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-160G18"  
 /clone\_11b="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:

FEATURES  
 source  
 1. 593  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-160G18"  
 /clone\_11b="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1:

ECORI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methyase. Size  
 selected DNA was cloned into the pBACE3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies).  
 BASE COUNT 88 a 159 c 213 g 133 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.51e-43 Length: 593  
 Score: 676.50 Matches: 136  
 Percent Similarity: 82.68% Conservative: 12  
 Best Local Similarity: 75.98% Mismatches: 30  
 Query Match: 60.03% Indels: 2  
 DB: 17 Gaps: 1

US-09-595-947C-10 (1-214) x AZ296526 (1-593)

Qy 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22  
 |||||  
 Db 547 CCTCATCCTTGGATGCGCTCACCATCCAGTGTCCCGAGACACACAACTTTTCCC 488  
 |||||  
 Qy 23 ArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThr 42  
 |||||  
 Db 487 GGAGCTCGGACCAAGTGTCTAGTTCATTCACCCCACTAGCCCACTCTCATA 428  
 |||||  
 Qy 43 ProGlyAsnCysAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLysLeuArg 62  
 |||||  
 Db 427 CTTAGGACTGCTCGAAGACAGAGTGGTGTGACTCCGAGGACCTCAGGAAGTCCGC 368  
 |||||  
 Qy 63 AlaArgArgGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArg 82  
 |||||  
 Db 367 GCGGACCGGAGGCGGCAAGAGCCCAAGAGTGGCTGCTCAGCAACAGCGCAAGA 308  
 |||||  
 Qy 83 SerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAla 102  
 |||||  
 Db 307 AGCCGCGCAAGAAGCCCAATGATCGGAGCGCAATCGCATGCACAACTCAACTCGGCG 248  
 |||||  
 Qy 103 LeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIle 122  
 |||||  
 Db 247 CTGGATGGCTGCGCGGTGCTCTGCCCACTTCGCGGATGAGCCAAACTTACAAAGATC 188  
 |||||  
 Qy 123 GluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAla 142  
 |||||  
 Db 187 GAGACCTGCGCTTCGCCCACTACATCTGGCACTGACTCAGACGCTGCGCATAGCG 128  
 |||||  
 Qy 143 AspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGlySerPro 162  
 |||||  
 Db 127 GACCACAGCTTCTATGCGCGGAGCCCTGTGCGCTG---GGAGAGCTGGGAGGCC 72  
 |||||  
 Qy 163 GlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlySer 181  
 |||||  
 Db 71 GGAGGGGTCCAAACGGGACTGGGCTCATCTACTCCCACTCTCCCAAGCGGAACC 15  
 |||||

RESULT 3  
 BQ169355 1098 bp mRNA linear EST 01-MAY-2002  
 LOCUS STR00303 segmentation stage cDNA library Danio rerio cDNA clone  
 DEFINITION CB260 5', similar to NEUROGENIN 1, mRNA sequence.  
 ACCESSION BQ169355  
 VERSION BQ169355.1 GI:20376783  
 KEYWORDS EST.  
 SOURCE zebrafish.  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 1 (bases 1 to 1098)  
 Loppin,B., Pflumio,S., Steffan,T., Heyer,V., Furthauer,M., Thisse  
 C. and Thisse,B.  
 Expression of the zebrafish genome during embryogenesis (2002)  
 Unpublished (2002)



```
QY 55 ArgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGlu 74
Db 226 -----GGCAACAGAAAGGAGCGCGCCAGCGCGCGCGCGC-----TGCAG 270
QY 75 LeuAlaLeuSerLysGlnArgArgSerArgLysLysAlaAsnAspArgGluArgAsn 94
Db 271 GCGGTGCGTGCAGTGTGAAAGAAAGACCGCGGCTGAAGCTAACGACCGGGAACGAAC 330
QY 95 ArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPhePro 114
Db 331 CGCATGCATAACCTGAACGACCGCTGGATGAGCTGCGCGGCTCTCGCGGCTTTCGG 390
QY 115 AspAspAlaLeuThrLysLysLeuThrLeuArgPheAlaHisAsnTyrIleTrpAla 134
Db 391 GACGAACAACCTGACCAAGATGAGACTCTGCGCTTTGCCACAACTACATCTGGGCT 450
QY 135 LeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAla-----LeuGluPro 151
Db 451 CTGTCCGAGACCATCGCATCGACAG-----CTCAGGCGGGGAGGCGCAACGACACC 504
QY 152 ProAla-----ProHisCys---GlyGluLeuGlySerProGlyGlyProGly 167
Db 505 CQTGCTGCTTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 564
QY 168 AspTyrGlySerLeuTyrSerProValSerGlnAlaGlySerLeuSerProAlaAlaSer 187
Db 565 AGCTGG-----AGCCCCAGGGATCTCTGCTGCGGCTCTCTCCCGCG----- 606
QY 188 LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySer 207
Db 607 -----TACTGCGCGTNCAGCCCGGCGCAGC 630

RESULT 5
BG808248 600 bp mRNA linear EST 20-DEC-2001
LOCUS 2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
DEFINITION mRNA sequence.
ACCESSION BG808248
VERSION BG808248.1 GI:17955225
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 600)
Molecular cloning of the mouse retina by EST sequencing
and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE 21671825
COMMENT Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3645
Fax: 713 790 0325
FEATURES
Location/Qualifiers
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Manniatidis); Cloning Technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGCTCAATCTGATG---. Other
information regarding entire library may be found at
http://pga.swmed.edu/Data/Libraries/microarray_cdna_librar
ies.htm."
BASE COUNT 94 a 238 c 161 g 107 t
```

```
ORIGIN
Alignment Scores: 6.67e-15 Length: 600
Pred. No.: 310.50 Matches: 73
Score: 58.39% Conservative: 7
Percent Similarity: 53.28% Mismatches: 35
Best Local Similarity: 27.55% Indels: 23
Query Match: 12 Gaps: 3
DB:
US-09-595-947C-10 (1-214) x BG808248 (1-600)
QY 91 ArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeu 110
Db 1 CGCAGGCGCAACCGCATGCACACCTAACACGCGCGCTGAGCGCGCTGCGGAGGTGCTG 60
QY 111 ProThrPheProAspAlaLysLeuThrLysLysLeuThrLysLysLeuThrLysLys 130
Db 61 CCCACCTTCCCGGAGGATGCCAAGCTCAGAAAGATCGAGACGCTGCGCTTCGCCCAAT 120
QY 131 TyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAla----- 148
Db 121 TACATCTGGGCGCTCACCAGAGACTCTGCGCTGGCGGACCACTGCCCGCGCGCGTGGC 180
QY 149 -----LeuGluProProAlaProHisCysGlyGlu 158
Db 181 CTCAGGGGCGCTCTTCACGAGGCGGTGCTCTCTGAGCGCGGAGCTGCGCTCGCGCC 240
QY 159 LeuGlySerProGlyGlyProGlyProGlyAspTrpGlySerLeuTyrSerProValSerGln 178
Db 241 AGCGGGGACAGCGCTTCTCCAGCTTCTCCTCGGAGCTGACCAACACCGCGCGCTCATCC 300
QY 179 AlaGlySerLeuSerPro-----AlaAlaSerLeuGluArgPro 192
Db 301 TCCACTCCAGCTCCCATACAGCTCAGCTTATCCCGCGCTAGCCCGCGGTACACGT 359
QY 193 GlyLeuGlyAlaThrSerSerAla-----CysLeuSerPro 205
Db 360 GGACTACTGGCAGCCCGCCACCTCCGGAGAGCATCGTTATGCGCTCACCT 410
RESULT 6
BE780690 1039 bp mRNA linear EST 20-OCT-2000
LOCUS 601469349Fl NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3872172 5',
DEFINITION mRNA sequence.
ACCESSION BE780690
VERSION BE780690.1 GI:10201888
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 1039)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Inceye Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9626 row: b column: 13
High quality sequence stop: 692.
Location/Qualifiers
1..1039
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3872172"
/clone_lib="NIH_MGC_67"
FEATURES
source
```

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE AUTHORS	TITLE	JOURNAL
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

## REFERENCE AUTHORS

**TITLE**

JOURNAL  
REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

## FEATURES

## ORIGIN

Pred. No.:  
Score:

Best Local  
Query Match

US-09-595-

Qy 321

533

Db 495

Qy 68

Qy 86 1

Db 399

QY 100

QY 126





```

QY 104 AspaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGlu 123
|||||
Db 484 GATTCCTGAGGAGGTTCTACCGCTATTACCGGAGCAGCCAACTACCAAGATAGAG 543
|||||

QY 124 ThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
|||||
Db 544 ACCTTGGCGCTTGGCCCAACTACATCTGGGCTCTTAGCGAAACTTTGGCG 594
|||||

RESULT 10
AL540071 804 bp mRNA linear EST 16-FEB-2001
LOCUS
DEFINITION
AL540071 LTI_FL013_FBRn1 Homo sapiens cDNA clone CS0DF035YD19 5
prime, mRNA sequence.
ACCESSION
AL540071
VERSION
AL540071.1 GI:12869886
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 804)
AUTHORS
Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF035YD19"
/clone_lib="LTI_FL013_FBRn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)."
/lab_host="DRI0B"
/notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 154 a 270 c 278 g 99 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 1,78e-11 Length: 804
Score: 268.50 Matches: 64
Percent Similarity: 51.01% Conservative: 12
Best Local Similarity: 42.95% Mismatches: 28
Query Match: 23.82% Indels: 45
DB: Gaps: 4

US-09-595-947C-10 (1-214) x AL540071 (1-804)

QY 31 CysProThrSerAlaProProSerProThrArgThrProGlyAsnCys----- 46
|||
Db 359 TGCTGCTCGGATCGGCGCCCGCTTGGCGCCCTGACCCCGCTCATCCAGCGCG 418
|||||

QY 47 -----AlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
|||||
Db 419 ACGAAGAGAGGAGGAGCGCGCGCTGAGGGCGGGCGCTCGGAGCGCGGGC 478
|||||

QY 61 LeuArgAlaArgGlyGlyArgSer----- 69
|||||
Db 479 TTA---NGCCGGCAANGGGCGCGCGCTGCTGGGGTGGCGAGGGCTGCCG 535
|||||

QY 70 -----ArgProLysSerGluLeuAla 76
|||||

```

```

Db 536 CCGCACGGCTGCTGGTCTGTGTACAGATTGCAACGGCGCCCTTCCCGGCGCGGCC 595
|||||
QY 77 LeuSerLysGln-----ArgArgSerArgArgLys 86
|||||
Db 596 GTCTCCCGAGGCGCCAAAGACGGCGAGCGGTGCGAGCATCAAGAAGACCGGTAGCTG 655
|||||
QY 87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106
|||||
Db 656 AAGCCCAACCAACCGGAGGAGGCGGATCGCAACCTCAACGCGGCGCTGGAGCGGCTG 715
|||||
QY 107 ArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeuArg 126
|||||
Db 716 CGCGAGGTGCTCCCGACGTTCCCGGAGGAGCGGCAAGCTCACCAGATCGAGACMCTGGCG 775
|||||
QY 127 PheAlaHisAsnTyrIleTrpAlaLeu 135
|||||
Db 776 TTGCCCCAACAACTACATCTGGGGACTC 802
|||||

RESULT 11
AL147434
LOCUS
DEFINITION
AL147434 401 bp mRNA linear EST 27-JUN-2001
XENOPUS_SOURCE_ID:xlnc001h24 5', similar to gb:gb|U67779.1|XL067779
Xenopus laevis neurogenin related 1b (XENOPUS);, mRNA sequence.
ACCESSION
AL147434
VERSION
AL147434.1 GI:6195330
KEYWORDS
EST.
SOURCE
African clawed frog.
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodidae; Xenopus.
REFERENCE
1 (bases 1 to 401)
AUTHORS
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, F., Underwood, K., Theising, B., Bowers, Y., Person
, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterson, R. and Wilson, R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: http://www.resgen.com/
Seq primer: -40RP from Gibco
High quality sequence stop: 264.
Location/Qualifiers
1..401
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XENOPUS_SOURCE_ID:xlnc001h24"
/clone_lib="Xenopus laevis oocyte"
/tissue_type="oocyte (stages 5 and 6)"
/lab_host="top-10 F'"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA.
EcoRI-XhoI cut cDNA was then ligated into UniLap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
end. SS-library phagemids were prepared by mass excision
from the original library and normalized by hybridization
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and

```

electroporated into Top-10 F'. Original library  
 contruction by Bruce Blumberg (Blumberg et al., 1991  
 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,  
 2923-2935). Normalized by Jihwan Song (Song, Cho and  
 Blumberg, unpublished). Note: This is a Xenopus Gene  
 Collection (XGC) library."

BASE COUNT 87 a 140 c 84 g 90 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9.97e-12 Length: 401  
 Score: 267.50 Matches: 62  
 Percent Similarity: 58.73% Conservative: 12  
 Best Local Similarity: 49.21% Mismatches: 35  
 Query Match: 23.74% Indels: 17  
 DB: 10 Gaps: 2

US-09-595-947C-10 (1-214) x AW147434 (1-401)

QY 82 ArgSerArgGlyLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101  
 DB 2 AAAACCCGCGGCTTAAGCCCAATACCGGAGAGAAATCGCATGCCACCTGAACTAT 61  
 QY 102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121  
 DB 62 GCGCTGATTCTCTGAGGAGGCTTACCGTCATTACCGAGAGCGCAAACTCACCAG 121  
 QY 122 IleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIle 141  
 DB 122 ATAGAGACCTTGGCTTTGCCCAACTACATCTGGGCTCTTAGCGAACTTTGGCGCTG 181  
 QY 142 AlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySer 161  
 DB 182 GCGGACCACTGACGAGTCTACTTCCACCCAGCA-----GCACCATATTGGTACAG 235  
 QY 162 ProGlyGlyProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlySer 181  
 DB 236 GACTCTATCT-----TCC 250  
 QY 182 LeuSerProAlaAlaSerLeuGluArgProGlyLeuLeuGlyAlaThrSerSerAla 201  
 DB 251 CTGAGCCCTCTGGAGCTGCAGCTCGTCCCATCTTCAACTCTTGGGACTCTCTCTCC 310  
 QY 202 CysLeuSerProGlySer 207  
 DB 311 TCGACCAAGCCCTGCCAGC 328

RESULT 12  
 AV673464 588 bp mRNA linear EST 05-OCT-2000  
 LOCUS  
 DEFINITION AV673464 Nori Satoh unpublished cDNA library Ciona intestinalis  
 CDNA clone c1b808 5', mRNA sequence.

ACCESSION AV673464  
 VERSION AV673464.1 GI:10111463  
 KEYWORDS EST.  
 SOURCE Ciona intestinalis.

ORGANISM Ciona intestinalis.

REFERENCE Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 Phlebobranchia; Clonidae; Ciona.

AUTHORS 1 (bases 1 to 588)  
 TITLE Sato, N., Satou, Y., Kohara, Y. and Shin-i, T.  
 JOURNAL Expressed genes in Ciona intestinalis

COMMENT Unpublished (2000)

CONTACT: Nori Satoh  
 Department of Zoology  
 Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoheascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1. .588

/organism="Ciona intestinalis"

FEATURES

source

/db\_xref="taxon:7719"  
 /clone="c1b808"  
 /clone\_lib="Nori Satoh unpublished cDNA library"  
 /tissue\_type="whole animal"  
 /dev\_stage="tailbud"  
 /note="vector: pBluescript SK"

BASE COUNT 205 a 146 c 117 g 120 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.96e-10 Length: 588  
 Score: 251.00 Matches: 51  
 Percent Similarity: 70.97% Conservative: 15  
 Best Local Similarity: 54.84% Mismatches: 15  
 Query Match: 22.27% Indels: 12  
 DB: 10 Gaps: 2

US-09-595-947C-10 (1-214) x AV673464 (1-588)

QY 48 GluAlaGluGluGlyCysArgGlyAlaProArgLysLeuArgAlaArgArgGlyGly 67  
 DB 14 GAGGAGAAAG-----CCAAAAAGAGCGCAGACGCGAA----- 49  
 QY 68 ArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLysLys 87  
 DB 50 -----AGAAAGTCAGTGACAATTACAAAGTAAAAAAGATTGACCATCAAAA 97  
 QY 88 AlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArg 107  
 DB 98 GCTAATGACAGAGAGAAACCGAATGATGGTCTTAATGACCGACTAGAAAGTAAGA 157  
 QY 108 GlyValLeuProThrPheProAspAlaLeuLysLeuThrLysIleGluThrLeuArgPhe 127  
 DB 158 CATGTTTTCGCGACCTACCCGACGAGACCAAAATTAACAAAAATCGAAACCTTGCATT 217  
 DB 218 GCATCAATATATTGTTGGTGTGTTGAGTGAATGTTAAAA 256

RESULT 13

AV995230

LOCUS

DEFINITION

AV995230

ACCESSION

AV995230

VERSION

AV995230.1

KEYWORDS

EST.

SOURCE

Ciona intestinalis.

ORGANISM

Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Clonidae; Ciona.

REFERENCE 1 (bases 1 to 490)

AUTHORS Sato, N., Satou, Y., Kohara, Y. and Shin-i, T.

TITLE Expressed genes in Ciona intestinalis

JOURNAL Unpublished (2000)

COMMENT Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoheascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1. .490

/organism="Ciona intestinalis"

/db\_xref="taxon:7719"

/clone="c1b40g24"

/clone\_lib="Nori Satoh unpublished cDNA library, tailbud

embryo"

/tissue\_type="whole animal"

/dev\_stage="tailbud embryo"

/note="Vector: pBluescript SK"

BASE COUNT 171 a 125 c 95 g 99 t

AV995230 490 bp mRNA linear EST 15-MAR-2002  
 AV995230 Nori Satoh unpublished cDNA library, tailbud embryo Ciona  
 intestinalis cDNA clone c1b40g24 5', mRNA sequence.

ACCESSION

AV995230

VERSION

AV995230.1

KEYWORDS

EST.

SOURCE

Ciona intestinalis.

ORGANISM

Ciona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Clonidae; Ciona.

REFERENCE 1 (bases 1 to 490)

AUTHORS Sato, N., Satou, Y., Kohara, Y. and Shin-i, T.

TITLE Expressed genes in Ciona intestinalis

JOURNAL Unpublished (2000)

COMMENT Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoheascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1. .490

/organism="Ciona intestinalis"

/db\_xref="taxon:7719"

/clone="c1b40g24"

/clone\_lib="Nori Satoh unpublished cDNA library, tailbud

embryo"

/tissue\_type="whole animal"

/dev\_stage="tailbud embryo"

/note="Vector: pBluescript SK"



**AUTHORS** Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.

**TITLE** Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 1003)  
**AUTHORS** Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

**TITLE** Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

**JOURNAL** Unpublished  
**REFERENCE** 3 (bases 1 to 1003)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (12-APR-2000)  
**COMMENT** This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

# FEATURES

**Source**  
1. .1003  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="141A20"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG141BAL01L1-end : T7"  
**BASE COUNT** 172 a 274 c 314 g 236 t 7 others  
**ORIGIN**

**Alignment Scores:**  
**Pred. No.:** 4.59e-09 **Length:** 1003  
**Score:** 239.00 **Matches:** 78  
**Percent Similarity:** 44.98% **Conservative:** 16  
**Best Local Similarity:** 37.32% **Mismatches:** 93  
**Query Match:** 21.21% **Indels:** 22  
**DB:** 17 **Gaps:** 7

US-09-595-947C-10 (1-214) x CNS02IB1 (1-1003)

QY 6 SerGlyAlaProThrValGlnValThrArg-GluThrGluArgSerPhePro----ArgAl 24  
Db 841 CTGGGGCGGAGRACAGCGACTCCGACGAGACAACTGAAGACGACGACGCGCGC 782  
QY 24 aSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGl 44  
Db 781 AGGACGACGTGGACGAGCGAGCTCGGCGCGCGCGCGACGACGACGACGACGCGG 722  
QY 44 yAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLysLeuArgAlaAr 64  
Db 721 GGGAGACGACGAGGACGACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 674  
QY 64 gArgGlyGlyArg-----SerArgProLysSerGluLeuAlaLe 77  
Db 673 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 614  
QY 77 uSerLysGluArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMethI 97  
Db 613 AATCGAGCGCTCCAAAGATGCGCGGATAAGGCCAACCGCGCGGAGGAGGAGGAGGAGGAGG 554  
QY 97 sAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAl 117  
Db 553 CGACCTGAACCTCCGCGCTCGCAATCTCGGTAAGATGGTCTCTTCTCTCTCTCTCTCTCTCT 494  
QY 117 aLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTrpIleTrpAlaLeuThrGl 137  
Db 493 GAACACTGTCCAAATCGACGACGCTGAGGTGGCCAAACATACATCTGGGCCCTCTGCGGA 434  
QY 137 nThrLeuArg-IleAlaAspHisSerLeuTyraLeuGluProProAlaProHisCysG 157  
Db 433 GATTTTACGCTCTGGGAAAGACCCGACCTGGTGT-----CCTACGTGCAGACGCTGTG 380

QY 157 lyCluLeuGlySerProGlyGlyProGlyAspTrpGlySerLeuTySerProVals 177  
Db 379 CAAGG---GCCTGTCCAGCCACCACCACTGGTGGCGGCTGCCGTGCAGCTCAACTC 323  
QY 177 erGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeu---LeuG 196  
Db 322 CCGTAACCTTCCTCACAGAGCAGCAGTGCC-----AGGACGGCGCGGTACGG 275  
QY 196 lyAlaThrSerSerAlaCysLeu 203  
Db 274 GCCCGGCTCCTTCTCCATGCATT 252

Search completed: April 9, 2003, 09:48:49  
Job time : 1767 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 11:37:10 ; Search time 4041 Seconds  
(without alignments)  
10514.748 Million cell updates/sec

Title: US-09-595-947C-1  
Perfect score: 1460  
Sequence: 1 gcaggtacgcagagagcag.....agagtgcctaatccagtgt 1460

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 8

Total number of hits satisfying chosen parameters: 3419321

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rod.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1460	100.0	1491	6	A91167 Sequence 1
2	1460	100.0	1491	10	Y10619 R.norvegicu
3	634	43.4	166677	2	AC127817 Rattus no
4	92	6.3	861	10	U76208 Mus muscu
5	92	6.3	1861	10	AF364300 Mus muscu
6	92	6.3	5567	10	Y09167 M.musculus
7	92	6.3	138070	2	AC109783 Mus muscu
8	92	6.3	185806	2	AC127417 Mus muscu
9	35	2.4	6123	9	AF303002 Homo sapi
10	35	2.4	91531	2	Continuation (4 of
11	35	2.4	179697	9	AC023886 Homo sapi
12	32	2.2	1330	9	AJ133776 Homo sapi
13	32	2.2	5340	9	AF234829 Homo sapi
14	32	2.2	165110	9	AL450311 Human DNA
15	32	2.2	173341	2	AC021954 Homo sapi
16	26	1.8	170896	2	AC011010 Homo sapi
17	25	1.7	25	6	A91170 Sequence 4
18	23	1.6	735	10	U67776 Mus musculu
19	23	1.6	770	5	AF123884 Gallus ga
20	23	1.6	790	5	AJ012659 Gallus ga
21	23	1.6	932	10	Y09166 M.musculus
22	23	1.6	1315	10	U63841 Mus musculu
23	23	1.6	1333	6	AR023715 Sequence
24	23	1.6	1341	5	AF109014 Gallus ga
25	23	1.6	1385	10	U76207 Mus musculu
26	23	1.6	1412	10	Y07621 M.musculus
27	23	1.6	1880	5	AF303000 Gallus ga
28	23	1.6	10393	10	AF303001 Mus muscu
29	23	1.6	71538	2	AC118243 Mus muscu
30	23	1.6	123855	2	AC102600 Mus muscu
31	23	1.6	174688	2	AC124395 Mus muscu
32	21	1.4	310	6	I84656 Sequence 12
33	21	1.4	1164	5	GGNEURODL
34	21	1.4	1268	6	AR023709 Sequence
35	21	1.4	1268	9	HSU63842 Human neuro
36	21	1.4	1352	6	I84655 Sequence 10
37	21	1.4	1535	6	AR023708 Sequence
38	21	1.4	1550	6	AR103242 Sequence
39	21	1.4	1550	6	BD000105 Different
40	21	1.4	1550	9	HSU58681 Homo sapien
41	21	1.4	1596	5	AF060885 Gallus ga
42	21	1.4	1675	9	BC008687 Homo sapi
43	21	1.4	1717	9	BC028226 Homo sapi
44	21	1.4	1831	10	MUSNDRF
45	21	1.4	1843	10	D82868 Rattus norv

# ALIGNMENTS

RESULT 1  
A91167  
LOCUS A91167  
DEFINITION Sequence 1 from Patent WO9827206.  
ACCESSION A91167  
VERSION A91167.1 GI:5740202  
KEYWORDS  
SOURCE Rattus sp.  
ORGANISM Rattus sp.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 1491)  
AUTHORS Icard-Liepkalns, C., Mallet, J. and Corresponding, N.A.  
JOURNAL Patent: WO 9827206-A 1 25-JUN-1998;

ICARD LIEPKALNS CHRISTINE (FR); MALLETT JACQUES (FR)

## FEATURES

Location/Qualifiers  
 1. .1491  
 /organism="Rattus sp."  
 /db\_xref="taxon:10118"  
 459..1103  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAB69414.1"  
 /db\_xref="GI:6740203"  
 DCEAEGACDRCGTSRLRRRGGRNRPKSELASKORRRRRKANDRENRHNLNSA  
 LDALRGVLPFPDDAKLTIETLRFANHYIWAULTQRLRIADHSFYGPPEPPVGEIGLS  
 PGGSGSDGMSIYSPVQASGSPSLASLEFPGLQVPSPLLPTLFLVSDPL"

BASE COUNT 307 a 487 c 413 g 284 t

ORIGIN  
 Query Match 100.0%; Score 1460; DB 6; Length 1491;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCAGGTAGCAGGAGGAGCAGTCCCTGGGCCCCCGCTTGGTGTGATGGCCCGTGGCAGAGGCA 60  
 DB 1 GCAGGTAGCAGGAGGAGCAGTCCCTGGGCCCCCGCTTGGTGTGATGGCCCGTGGCAGAGGCA 60  
 QY 61 GCAGCCCGGAGGACGCTCCTGTGTCCGGGAGAGCAGATAAAGCGTGCCAGGGGACACA 120  
 DB 61 GCAGCCCGGAGGACGCTCCTGTGTCCGGGAGAGCAGATAAAGCGTGCCAGGGGACACA 120  
 QY 121 CGATTAGCAGTCAAGAGTCCCTCTGGGTCTCACCCTGCACAGAGCGGAGGACCCCT 180  
 DB 121 CGATTAGCAGTCAAGAGTCCCTCTGGGTCTCACCCTGCACAGAGCGGAGGACCCCT 180  
 QY 181 CCGAGCTCTTTGTGCTCCAGACGCAATTTACTCCAGGCGAGGCGCTTCAGCTCAG 240  
 DB 181 CCGAGCTCTTTGTGCTCCAGACGCAATTTACTCCAGGCGAGGCGCTTCAGCTCAG 240  
 QY 241 CAAAGCTCGNAGCGAGGAGGTTGAGTATCCACCGCTGCTTACTCTGACCTGACACCC 300  
 DB 241 CAAAGCTCGNAGCGAGGAGGTTGAGTATCCACCGCTGCTTACTCTGACCTGACACCC 300  
 QY 301 GCAGCTCTGTTCTTTGAGCCCGGAGTAACTAGGTAAACATTTAGGAACCTCCAAAGGG 360  
 DB 301 GCAGCTCTGTTCTTTGAGCCCGGAGTAACTAGGTAAACATTTAGGAACCTCCAAAGGG 360  
 QY 361 TAGAAGAGGAGGTGGTGGGCGTACTCTAGTCCCGGTGGAGTGACCTTAAGTCAGAG 420  
 DB 361 TAGAAGAGGAGGTGGTGGGCGTACTCTAGTCCCGGTGGAGTGACCTTAAGTCAGAG 420  
 QY 421 ACTGTACACCCCTCCATTTTCCCAACCTTCAGATGGGCGCTCATCCCTTGGATG 480  
 DB 421 ACTGTACACCCCTCCATTTTCCCAACCTTCAGATGGGCGCTCATCCCTTGGATG 480  
 QY 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCGGAGGCTCGGACCAAG 540  
 DB 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCGGAGGCTCGGACCAAG 540  
 QY 541 AAGTGTCTAGTTCOAATTCACCCCACTAGCCCACTCTCGTACCGAGGAGTGTCTCCG 600  
 DB 541 AAGTGTCTAGTTCOAATTCACCCCACTAGCCCACTCTCGTACCGAGGAGTGTCTCCG 600  
 QY 601 AAGCAGAGGAGGTGACTCGGAGGAGACATCGAGAGAGTCCGTCGCGGCGGAGGGC 660  
 DB 601 AAGCAGAGGAGGTGACTCGGAGGAGACATCGAGAGAGTCCGTCGCGGCGGAGGGC 660  
 QY 661 GCACAGGCGCCCAAGAGGAGTGTGCACTGAGCAAGCAGCAGCAAGCGCGCAAGAGG 720  
 DB 661 GCACAGGCGCCCAAGAGGAGTGTGCACTGAGCAAGCAGCAGCAAGCGCGCAAGAGG 720  
 QY 721 CCAACGACCGGAGCGCAACCGCATGACCAACCTTAATCTCCGCGCTGGATGCGCTGCGG 780  
 DB 721 CCAACGACCGGAGCGCAACCGCATGACCAACCTTAATCTCCGCGCTGGATGCGCTGCGG 780

## RESULT 2

## RNRELAXT

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

1491 bp mRNA linear ROD 06-MAY-1997  
 R. norvegicus mRNA for transcriptional regulator, Relax.

Y10619  
 Y10619.1 GI:2072737

Relax; transcriptional regulator.  
 Rattus norvegicus

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 1491)  
 Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.

Relax, a novel rat bHLH transcriptional regulator transiently  
 expressed in the ventricular proliferating zone of the developing

central nervous system

J. Neurosci. Res. 48 (2), 146-158 (1997)

97276390

9130143

2 (bases 1 to 1491)

Ravassard, P.

Direct Submission

Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,

Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013 Paris, FRANCE

FEATURES  
source Location/Qualifiers

1..1491  
/organism="Rattus norvegicus"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
/clone\_lib="RT-PCR"  
459..1103  
/function="transcriptional regulator"  
/codon\_start=1  
/product="Relax"  
/protein\_id="CAA71630.1"  
/db\_xref="GI:2072738"  
/db\_xref="SPTREMBL:O08718"  
/translation="NAPFLDAPTIQVSQTOOPFGASDHEVLSSNTPSPPTLVPR  
DCSEAEAGCGTSSRLKRRRNPKESEALSQRSSRRKANDRRNRNHLNSA  
LDALRGVLPPTDDAKLKIETLRPHNY IVALTQTLRIADHSFYGPPEVPCGELGS  
PGGSSGDWGSYISPVQAGSLSPASLEEFPLQVPSPSCLLPGLTFVFSDFL"

CDS  
307 a 487 c 413 g 284 t

BASE COUNT  
ORIGIN

Query Match 100.0%; Score 1460; DB 10; Length 1491;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCGCGTTCGTGATTGGCCCGGTGGCAGCAGGCA 60  
|||||  
Db 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCGCGTTCGTGATTGGCCCGGTGGCAGCAGGCA 60  
|||||

QY 61 GCAGCCCGCAGCAGCGTCTGGTCCGGCAGCAGATAGAGCGTGCAGGGGACACA 120  
|||||  
Db 61 GCAGCCCGCAGCAGCGTCTGGTCCGGCAGCAGATAGAGCGTGCAGGGGACACA 120  
|||||

QY 121 CGATTAGCAGCTCAGAAAGTCCCTCTGGGCTCACCACTGCACAGAGCGCGAGACCCCT 180  
|||||  
Db 121 CGATTAGCAGCTCAGAAAGTCCCTCTGGGCTCACCACTGCACAGAGCGCGAGACCCCT 180  
|||||

QY 181 CCGAGTCTCTGTTCTTCTGCTCCAGACGCAATTTACTCCAGCGAGGCGGCTGCAGTCTCAG 240  
|||||  
Db 181 CCGAGTCTCTGTTCTTCTGCTCCAGACGCAATTTACTCCAGCGAGGCGGCTGCAGTCTCAG 240  
|||||

QY 241 CAAAGCTTCCAGCAGCAGAGGGTTCAGCTATCCACCGCTCTGACTCTGACCTGACCAACC 300  
|||||  
Db 241 CAAAGCTTCCAGCAGCAGAGGGTTCAGCTATCCACCGCTCTGACTCTGACCTGACCAACC 300  
|||||

QY 301 CGAGCTCTCTGTTCTTTTGGAGCCCGGAGTAACCTAGTAACTATTAGGAACCTCCAAAGG 360  
|||||  
Db 301 CGAGCTCTCTGTTCTTTTGGAGCCCGGAGTAACCTAGTAACTATTAGGAACCTCCAAAGG 360  
|||||

QY 361 TAGAAGAGGGAGTGGGTGGGTACTCTAGTCCCGTGGAGTGACCTTAAGTCAGAG 420  
|||||  
Db 361 TAGAAGAGGGAGTGGGTGGGTACTCTAGTCCCGTGGAGTGACCTTAAGTCAGAG 420  
|||||

QY 421 ACTGTCCACACCCCTTCCATTTTCCCAACCTCAGGATGGCGCTCATCCCTTGATG 480  
|||||  
Db 421 ACTGTCCACACCCCTTCCATTTTCCCAACCTCAGGATGGCGCTCATCCCTTGATG 480  
|||||

QY 481 CGCCCACTCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGGCTCCGAGCAGC 540  
|||||  
Db 481 CGCCCACTCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGGCTCCGAGCAGC 540  
|||||

QY 541 AAGTGTCTAGTTCCTCAATTCACCCACTAGCCCACTCTCTGTAACCGAGGACTGTCTCG 600  
|||||  
Db 541 AAGTGTCTAGTTCCTCAATTCACCCACTAGCCCACTCTCTGTAACCGAGGACTGTCTCG 600  
|||||

QY 601 AAGCAGAGCAGTGTCTCCGAGGACATCGAGGAGCTCCGTGCGCGCGGAGGCG 660  
|||||  
Db 601 AAGCAGAGCAGTGTCTCCGAGGACATCGAGGAGCTCCGTGCGCGCGGAGGCG 660  
|||||

QY 661 GCAACAGGCCAAGAGCGAGTTGGCACTGAGCAAGCAGGACGACGAAGCCGCGCAAGAG 720  
|||||  
Db 661 GCAACAGGCCAAGAGCGAGTTGGCACTGAGCAAGCAGGACGACGAAGCCGCGCAAGAG 720  
|||||

QY 721 CCAACGACGGGAGCGCAACCGCATGCAACAACTTAACCTCCGGCTGGATGCGCTGGCG 780  
|||||  
Db 721 CCAACGACGGGAGCGCAACCGCATGCAACAACTTAACCTCCGGCTGGATGCGCTGGCG 780  
|||||

QY 781 GTGTCTGCTCCACCTTCCCGGATGAGCCCAAACTTACAAGATCGAGACCTCGCTTCG 840  
|||||  
Db 781 GTGTCTGCTCCACCTTCCCGGATGAGCCCAAACTTACAAGATCGAGACCTCGCTTCG 840  
|||||

QY 841 CCCAACAATACATTTGGGACACTGACTCAGACGCTGGCATAGGGACACAGCTTCTACG 900  
|||||  
Db 841 CCCAACAATACATTTGGGACACTGACTCAGACGCTGGCATAGGGACACAGCTTCTACG 900  
|||||

QY 901 GCCCGAGGCGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGCG 960  
|||||  
Db 901 GCCCGAGGCGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGCG 960  
|||||

QY 961 ACTGGGCTCTATCTACTCTCCCACTTCCCAAGCTGTAGCTAGCCCTCAGCCACAGCTCAT 1020  
|||||  
Db 961 ACTGGGCTCTATCTACTCTCCCACTTCCCAAGCTGTAGCTAGCCCTCAGCCACAGCTCAT 1020  
|||||

QY 1021 TGGAGGAGTTCCTTGGCTGCGAGTGCAGCTCCCACTCCCTCTCTGTCTCCCGGACCC 1080  
|||||  
Db 1021 TGGAGGAGTTCCTTGGCTGCGAGTGCAGCTCCCACTCCCTCTCTGTCTCCCGGACCC 1080  
|||||

QY 1081 TGGTGTCTCAGACTTCTTGTGAAGGCGCCCAACAGCGCCCTGGCGGTGGCGCTGGCAG 1140  
|||||  
Db 1081 TGGTGTCTCAGACTTCTTGTGAAGGCGCCCAACAGCGCCCTGGCGGTGGCGCTGGCAG 1140  
|||||

QY 1141 AAGGAGGAGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCAGCATCTCGC 1200  
|||||  
Db 1141 AAGGAGGAGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCAGCATCTCGC 1200  
|||||

QY 1201 CCCTTCTGCTTTCATTAGTCAGGTCCTGATTAAACAGGATTCGCACAGTTCTCTGCT 1260  
|||||  
Db 1201 CCCTTCTGCTTTCATTAGTCAGGTCCTGATTAAACAGGATTCGCACAGTTCTCTGCT 1260  
|||||

QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCTCAGTGTGGCC 1320  
|||||  
Db 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAAACCTCTCAGTGTGGCC 1320  
|||||

QY 1321 ACCTCAAACTCCCGCTCCCAAGCAGAGAGCGTAGCAGTAAATAGTTGGGAGACTCCC 1380  
|||||  
Db 1321 ACCTCAAACTCCCGCTCCCAAGCAGAGAGCGTAGCAGTAAATAGTTGGGAGACTCCC 1380  
|||||

QY 1381 ATACTTCTGCTGACTCCCGCTCTTCAAACTCTGGGCGCTCCCAACCCAGCTTCTCC 1440  
|||||  
Db 1381 ATACTTCTGCTGACTCCCGCTCTTCAAACTCTGGGCGCTCCCAACCCAGCTTCTCC 1440  
|||||

QY 1441 AGAGTGACCTTAATCCAGTGT 1460  
|||||  
Db 1441 AGAGTGACCTTAATCCAGTGT 1460  
|||||

RESULT 3  
AC127817 166677 bp DNA linear HTG 19-JUL-2002  
LOCUS Rattus norvegicus clone CH230-259G16, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 80 unordered pieces.  
AC127817  
AC127817.1 GI:21908200  
HTG: HTGS\_PHASE1.  
Rattus norvegicus.  
Rattus norvegicus.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 166677)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshan, F.R., Allen, C.,  
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,



Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Homs, J.F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
 Oragunye, N., Oviato, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williams, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 166677)  
 Worley, K.C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GZXS  
 Center clone name: CH230-259G16  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 73194 bases at least Q40  
 Consensus quality: 79767 bases at least Q30  
 Consensus quality: 84278 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 80 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1032: contig of 1032 bp in length  
 \* 1033 1132: gap of unknown length  
 \* 1133 2146: contig of 1014 bp in length  
 \* 2147 2246: gap of unknown length

2247 \*  
 3350 \*  
 3650 \*  
 4696 \*  
 4796 \*  
 6192 \*  
 6193 \*  
 6292 \*  
 6968 \*  
 8068 \*  
 9287 \*  
 9387 \*  
 10393 \*  
 10493 \*  
 11602 \*  
 11703 \*  
 13173 \*  
 13174 \*  
 13274 \*  
 14295 \*  
 14395 \*  
 15983 \*  
 16083 \*  
 17116 \*  
 17216 \*  
 18362 \*  
 18462 \*  
 20372 \*  
 20472 \*  
 21519 \*  
 21619 \*  
 22638 \*  
 22738 \*  
 22739 \*  
 23988 \*  
 24089 \*  
 25266 \*  
 25366 \*  
 26676 \*  
 28415 \*  
 28515 \*  
 29921 \*  
 29922 \*  
 30022 \*  
 31951 \*  
 32051 \*  
 32052 \*  
 33981 \*  
 33981 \*  
 34081 \*  
 35825 \*  
 35925 \*  
 37572 \*  
 37573 \*  
 39232 \*  
 39233 \*  
 39333 \*  
 41121 \*  
 41221 \*  
 42662 \*  
 42663 \*  
 44648 \*  
 44748 \*  
 46212 \*  
 46312 \*  
 48275 \*  
 48376 \*  
 49451 \*  
 49550 \*  
 51299 \*  
 51300 \*  
 51300 \*  
 52749 \*  
 52849 \*  
 52850 \*  
 54511 \*  
 54611 \*  
 54612 \*  
 55929 \*  
 56029 \*  
 57531 \*  
 57532 \*  
 59080 \*  
 3549: contig of 1303 bp in length  
 3649: gap of unknown length  
 4695: contig of 1046 bp in length  
 4795: gap of unknown length  
 6192: contig of 1397 bp in length  
 6292: gap of unknown length  
 6967: contig of 1675 bp in length  
 8067: gap of unknown length  
 9286: contig of 1219 bp in length  
 9386: gap of unknown length  
 10392: contig of 1006 bp in length  
 10492: gap of unknown length  
 11602: contig of 1110 bp in length  
 11702: gap of unknown length  
 13173: contig of 1471 bp in length  
 13273: gap of unknown length  
 14294: contig of 1021 bp in length  
 14394: gap of unknown length  
 15982: contig of 1588 bp in length  
 16082: gap of unknown length  
 17115: contig of 1033 bp in length  
 17215: gap of unknown length  
 18361: contig of 1146 bp in length  
 18461: gap of unknown length  
 20371: contig of 1910 bp in length  
 20471: gap of unknown length  
 21519: contig of 1048 bp in length  
 21619: gap of unknown length  
 22638: contig of 1019 bp in length  
 22738: gap of unknown length  
 23988: contig of 1250 bp in length  
 24088: gap of unknown length  
 25265: contig of 1177 bp in length  
 25365: gap of unknown length  
 26675: contig of 1310 bp in length  
 26775: gap of unknown length  
 28414: contig of 1639 bp in length  
 28514: gap of unknown length  
 29921: contig of 1407 bp in length  
 30021: gap of unknown length  
 31951: contig of 1930 bp in length  
 32051: gap of unknown length  
 33980: contig of 1929 bp in length  
 34080: gap of unknown length  
 35824: contig of 1744 bp in length  
 35924: gap of unknown length  
 37572: contig of 1648 bp in length  
 37672: gap of unknown length  
 39232: contig of 1560 bp in length  
 39332: gap of unknown length  
 41120: contig of 1788 bp in length  
 41220: gap of unknown length  
 42662: contig of 1442 bp in length  
 42762: gap of unknown length  
 44647: gap of unknown length  
 46211: contig of 1464 bp in length  
 46311: gap of unknown length  
 48275: contig of 1964 bp in length  
 48375: gap of unknown length  
 49450: contig of 1075 bp in length  
 49550: gap of unknown length  
 51299: contig of 1749 bp in length  
 51300: gap of unknown length  
 52749: contig of 1350 bp in length  
 52849: gap of unknown length  
 54511: contig of 1662 bp in length  
 54611: gap of unknown length  
 54612: contig of 1317 bp in length  
 55928: gap of unknown length  
 56028: gap of unknown length  
 57531: contig of 1503 bp in length  
 57532: gap of unknown length  
 59080: contig of 1449 bp in length

```

* 59081 59180: gap of unknown length
* 59181 60513: contig of 1333 bp in length
* 60514 60613: gap of unknown length
* 60614 62621: contig of 2008 bp in length
* 62622 62721: gap of unknown length
* 62722 64707: contig of 1986 bp in length
* 64708 64807: gap of unknown length
* 64808 66176: contig of 1369 bp in length
* 66177 66276: gap of unknown length
* 66277 68684: contig of 2408 bp in length
* 68685 68784: gap of unknown length
* 68785 70841: contig of 2057 bp in length
* 70842 72278: gap of unknown length
* 72279 72378: contig of 1337 bp in length
* 72379 73868: gap of unknown length
* 73869 73968: contig of 1490 bp in length
* 73969 75979: contig of 2011 bp in length
* 75980 76079: gap of unknown length
* 76080 78068: contig of 1989 bp in length
* 78069 78168: gap of unknown length
* 78169 79454: contig of 1286 bp in length
* 79455 79554: gap of unknown length
* 79555 80709: contig of 1155 bp in length
* 80710 80809: gap of unknown length
* 80810 82754: contig of 1945 bp in length
* 82755 82854: gap of unknown length
* 82855 83396: contig of 2542 bp in length
* 83397 85496: gap of unknown length
* 85497 88065: contig of 2569 bp in length
* 88066 88165: gap of unknown length
* 88166 90986: contig of 2831 bp in length
* 90987 91096: gap of unknown length
* 91097 92577: contig of 1481 bp in length
* 92578 94520: contig of 1843 bp in length
* 94521 94620: gap of unknown length

Query Match
Best Local Similarity 100.0%; Score 634; DB 2; Length 166677;
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 827 GACCTCGCTTCGCCACAACTACATTTGGCACTGACTCAGACGCTGCGCATAGCGGA 886
DB 57641 GACCTCGCTTCGCCACAACTACATTTGGCACTGACTCAGACGCTGCGCATAGCGGA 57700

QY 887 CCACAGCTTACGGCCCGAGCCCTGTCGCTGCGGAGCTGGGAAGCCCGGAGG 946
DB 57701 CCACAGCTTACGGCCCGAGCCCTGTCGCTGCGGAGCTGGGAAGCCCGGAGG 57760

QY 947 GGGCTCAGGGGACCTGGGCTCTATCTACTCCCAAGTTTCCCAAGCTGGTAGCCTGAG 1006
DB 57761 GGGCTCAGGGGACCTGGGCTCTATCTACTCCCAAGTTTCCCAAGCTGGTAGCCTGAG 57820

QY 1007 CCCACAGCCTCATTGGAGGAGTTCCTGGCTGCGCTCAGGTGCCAGCTCCCATCTGCT 1066
DB 57821 CCCACAGCCTCATTGGAGGAGTTCCTGGCTGCGCTCAGGTGCCAGCTCCCATCTGCT 57880

QY 1067 GCTCCCGGGACCCCTGCTGCTTCAGACTTCTTGTGAAGGCCCAACAGCCCTGGGCG 1126
DB 57881 GCTCCCGGGACCCCTGCTGCTTCAGACTTCTTGTGAAGGCCCAACAGCCCTGGGCG 57940

QY 1127 GTGGGCGCTGCCAAGAGGAGGAGTACAGCTGTCTGAATGAAGGTAGTAGGCGCA 1186
DB 57941 GTGGGCGCTGCCAAGAGGAGGAGTACAGCTGTCTGAATGAAGGTAGTAGGCGCA 58000

QY 1187 CTCGAGCATCTCGCCCTTCCTGGCTTCATTAGTCAGTCCCTGATTTAACAGGATTCG 1246
DB 58001 CTCGAGCATCTCGCCCTTCCTGGCTTCATTAGTCAGTCCCTGATTTAACAGGATTCG 58060

QY 1247 CACAGTTCCTTGCTGTGCTGTCACAAAGGACATTCAGGCTGATCTCTTAAACC 1306
DB 58061 CACAGTTCCTTGCTGTGCTGTCACAAAGGACATTCAGGCTGATCTCTTAAACC 58120

```

```

QY 1307 TCCTCAGTGTGGCCACCTCAAACTCCCGCTCCRAGCAGAGAGCGGTAGCAATA 1366
DB 58121 TCCTCAGTGTGGCCACCTCAAACTCCCGCTCCRAGCAGAGAGCGGTAGCAATA 58180

QY 1367 GTTGGGAGACTCCCATACTTCCTGGTGACTCCGCGCTCTTTCAAACTCTCGGGGCTCCAA 1426
DB 58181 GTTGGGAGACTCCCATACTTCCTGGTGACTCCGCGCTCTTTCAAACTCTCGGGGCTCCAA 58240

QY 1427 CCACCGCTTTCTCCAGAGTACCTTAATCCAGTGT 1460
DB 58241 CCACCGCTTTCTCCAGAGTACCTTAATCCAGTGT 58274

RESULT 4
MMU76208
LOCUS
DEFINITION Mus musculus neurogenin 3 (ngn3) gene, complete cds.
ACCESSION U76208
VERSION U76208.1 GI:1815654
KEYWORDS
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 861)
AUTHORS Sommer,L., Ma,Q. and Anderson,D.J.
TITLE neurogenins, a novel family of atonal-related bHLH transcription
factors, are putative mammalian neuronal determination genes that
reveal progenitor cell heterogeneity in the developing CNS and PNS
Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
JOURNAL
MEDLINE 97153565
PUBMED 9000438
REFERENCE 2 (bases 1 to 861)
AUTHORS Sommer,L., Ma,Q. and Anderson,D.J.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1996) Biology 216-76, California Institute of
Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
USA
COMMENT On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES
source
1..861
/organism="Mus musculus"
/db_xref="taxon:10090"
160..804
/gene="ngn3"
160..804
/gene="ngn3"
/function="bHLH transcription factor"
/notes="NGN3"
/codon_start=1
/product="neurogenin 3"
/protein_id="AAC53029.1"
/db_xref="GI:1666912"
DCSEAEEVDCRGTSRLARRGRNRKSELALSKQRSRKRRKANDRRMRHNLNSA
LDALRGVLPFPDDAKLTKETLRFHNYIWTQTLRIADHSFYGPPEVPVCGELGS
PGSGNGDWGSIYSPVSOAGNLSPTASEEFGLOVPSFYLPLGALVFSDFL"

BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGGATGCGCTGCGGGTGTCTGCCACCTTCGCGGATGACGCCAACTACAAG 821
DB 463 GCGCTGGATGCGCTGCGGGTGTCTGCCACCTTCGCGGATGACGCCAACTACAAG 522

QY 822 ATCAGAGCCCTGCGCTTCGCCCACTACAT 853
DB 523 ATCAGAGCCCTGCGCTTCGCCCACTACAT 554

```

```

RESULT 5
AF364300
LOCUS Mus musculus neurogenin 3 gene, complete cds. linear ROD 03-MAY-2001
DEFINITION AF364300
ACCESSION AF364300
VERSION AF364300.1 GI:13937128
KEYWORDS Mus musculus.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1861)
AUTHORS Lee, J., Smith, S., Watada, H., Lin, J., Scheel, D., Wang, J., Mirmira, R.
and German, M.
TITLE Regulation of the pancreatic pro-endocrine gene neurogenin3
JOURNAL Diabetes (2001) In press
REFERENCE 2 (bases 1 to 1861)
AUTHORS Schwitzgebel, V. and German, M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2001) Hormone Research Institute, University of
California San Francisco, 513 Parnassus Ave, HSW1090, San
Francisco, CA 94145-0534, USA
FEATURES
source
1..1861
Location/Qualifiers
1..1861
/organism="Mus musculus"
/db_xref="taxon:10090"
<1093..>1737
/product="neurogenin 3"
1093..1737
/codon_start=1
/product="neurogenin 3"
/db_xref="GI:13937128"
DCSAEYVDCRGTSRKLRARGGRNRPKSELALSKQRSSRRKKANDRRNRHNLNSA
LDALRGVLPFPDDAKLTKIETLFAHNYIWAQTOTLRADHSFYGPPEPPVPCGELGS
PGGSGNDGWSIYSPVQAGNLSPASLEEFPGQVPPSSPYLLPGALVFDFL"
BASE COUNT 397 a 560 c 537 g 367 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 1861;
Best Local Similarity 100.0%; Pred. No. 1.5e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 762 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAG 821
Db 1396 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAG 1455
Qy 822 ATCGAGACCCCTGCGCTTCCGCCACAACTACAT 853
Db 1456 ATCGAGACCCCTGCGCTTCCGCCACAACTACAT 1487
RESULT 6
MMATH4B
LOCUS M.musculus MATH4B gene. 5567 bp DNA linear ROD 25-OCT-2000
DEFINITION Y09167
ACCESSION Y09167
VERSION Y09167.2 GI:11065737
KEYWORDS bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 5567)
AUTHORS Cau, E., Gradwohl, G., Fode, C. and Guilleminot, F.
TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron
progenitors
JOURNAL Development 124 (8), 1611-1621 (1997)
MEDLINE 97261963
PUBMED 9108377
REFERENCE 2 (bases 1 to 5567)

```

```

AUTHORS Jacquemin, P., Durvieux, S.M., Jensen, J., Godfraind, C., Gradwohl, G.,
Guilleminot, F., Madsen, O.D., Carmeliet, P., Dewerchin, M., Collen, D.,
Rousseau, G.G. and Lemaigre, F.P.
TITLE Transcription factor hepatocyte nuclear factor 6 regulates
pancreatic endocrine cell differentiation and controls expression
of the proendocrine gene ngn3
JOURNAL Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
MEDLINE 20285449
PUBMED 10825208
REFERENCE 3 (bases 1 to 5567)
AUTHORS Gradwohl, G.J.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,
F-67404 ILLKIRCH cedex, FRANCE
REMARK Revised by [4]
REFERENCE 4 (bases 1 to 5567)
AUTHORS Lemaigre, F.P.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
Research Unit, Louvain University Medical School, Avenue Hippocrate
75, box 7529, Brussels 1200, BELGIUM
COMMENT On Oct 31, 2000 this sequence version replaced gi:1666087.
FEATURES
Location/Qualifiers
1..5567
/organism="Mus musculus"
/strain="SV129"
/db_xref="taxon:10090"
/clone="lambda GEM12 clone 19C"
/cell_line="D3"
/cell_type="embryonic stem cells"
/clone_lib="Genomic library No.62, MboI partial digests
cloned in BamHI"
/dev_stage="embryos"
958
/bound_moiety="HNF6"
misc_binding 3692
/bound_moiety="HNF6"
misc_binding 4145..5567
gene /gene="MATH4B"
4145..4151
TATA_signal /gene="MATH4B"
/note="putative"
4923..5567
CDS /gene="MATH4B"
/codon_start=1
/product="neurogenin 3"
/protein_id="CAA70366.1"
/db_xref="GI:1666088"
/db_xref="SWISS-PROT:P70661"
/translation="MAPHLDALTIQVSPETQPPPGASDHEVLSNSTPPSPPTLIPR
DCSAEYVDCRGTSRKLRARGGRNRPKSELALSKQRSSRRKKANDRRNRHNLNSA
LDALRGVLPFPDDAKLTKIETLFAHNYIWAQTOTLRADHSFYGPPEPPVPCGELGS
PGGSGNDGWSIYSPVQAGNLSPASLEEFPGQVPPSSPYLLPGALVFDFL"
BASE COUNT 1271 a 1549 c 1564 g 1183 t
ORIGIN
Query Match 6.3%; Score 92; DB 10; Length 5567;
Best Local Similarity 100.0%; Pred. No. 1.4e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 762 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAG 821
Db 5226 GCGCTGGATGCGCTCGCGGGTGCTCTGCCACCTTCCCGGATGACGCCAACTTACAAG 5285
Qy 822 ATCGAGACCCCTGCGCTTCCGCCACAACTACAT 853
Db 5286 ATCGAGACCCCTGCGCTTCCGCCACAACTACAT 5317
RESULT 7
AC109783/c AC109783 138070 bp DNA linear HTG 07-FEB-2002
LOCUS

```

```

DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
unordered pieces,
AC109783.1 GI:18581594
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus.
SOURCE ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 138070)
McCombie,W.R., Baker,J.P., Ballja,V., Dedhia,N.N., de la
Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
Zucavern,R.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 138070)
McCombie,W.R.
Direct Submission
Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
----- Genome Center -----
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Project name: RP23-121F10
Clone name: RP23-121F10
Insert size: 173000; agarose-fp
Insert size: 141616; sum-of-contigs
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 3.70 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 17785: contig of 17785 bp in length
* 17786 17874: gap of unknown length
* 17875 30356: contig of 12482 bp in length
* 30357 30444: gap of unknown length
* 30445 42306: contig of 11862 bp in length
* 42307 42394: gap of unknown length
* 42395 53598: contig of 11204 bp in length
* 53599 53686: gap of unknown length
* 53687 64356: contig of 10669 bp in length
* 64356 74016: contig of 9573 bp in length
* 74017 83366: contig of 9262 bp in length
* 83367 83454: gap of unknown length
* 83455 92433: contig of 8901 bp in length
* 92433 92444: gap of unknown length
* 92444 100821: contig of 8378 bp in length
* 100822 100909: gap of unknown length
* 100910 107529: contig of 6620 bp in length
* 107530 107617: gap of unknown length
* 107618 114066: contig of 6449 bp in length
* 114067 114154: gap of unknown length
* 114155 118873: contig of 4719 bp in length
* 118874 118961: gap of unknown length
* 118962 123619: contig of 4638 bp in length
* 123620 123707: gap of unknown length
* 123708 128240: contig of 4533 bp in length
* 128241 128328: gap of unknown length
*
* 128329 132682: contig of 4354 bp in length
* 132683 132770: gap of unknown length
* 132771 136341: contig of 3571 bp in length
* 136342 136429: gap of unknown length
* 136430 138070: contig of 1641 bp in length.
Location/Qualifiers
1. 138070
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-121F10"
BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN
Query Match 6.3% Score 92; DB 2: Length 138070;
Best Local Similarity 100.0%; Pred. No. 1.2e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 762 GCGCTGGATGCGCTGCGGGTGCTGCTGCCACCTTCCCGGATGAGCCCAACTTACAAG 821
Db 110917 GCGCTGGATGCGCTGCGGGTGCTGCTGCCACCTTCCCGGATGAGCCCAACTTACAAG 110858
QY 822 ATCGAGACCTCGGCTTCCGCCACACTACAT 853
Db 110857 ATCGAGACCTCGGCTTCCGCCACACTACAT 110826
-----
RESULT 8
AC127417 185806 bp. DNA linear HTG 15-JUL-2002
LOCUS Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
DEFINITION SEQUENCE, 36 unordered pieces.
AC127417
AC127417.1 GI:21759524
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 185806)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185806)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0459M02
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173518 bases at least Q40
Consensus quality: 177276 bases at least Q30
Consensus quality: 179611 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 184535; sum-of-contigs
Quality coverage: 3.66 in Q20 bases; agarose-fp
Quality coverage: 3.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1042: contig of 1042 bp in length
* 1043
* 1142: gap of unknown length
* 1143
* 2216: contig of 1074 bp in length
* 2217
* 3838: contig of 1522 bp in length
* 3839
* 3938: gap of unknown length
* 3939
* 5396: contig of 1458 bp in length
* 5397
* 5496: gap of unknown length
* 5497
* 6874: contig of 1378 bp in length
* 6875
* 6974: gap of unknown length
* 6975
* 8481: contig of 1507 bp in length
* 8482
* 8581: gap of unknown length
* 8582
* 9874: contig of 1293 bp in length
* 9875
* 9974: gap of unknown length
* 9975
* 11483: contig of 1509 bp in length
* 11484
* 11583: gap of unknown length
* 11584
* 13834: contig of 2251 bp in length
* 13835
* 13934: gap of unknown length
* 13935
* 16035: contig of 2101 bp in length
* 16036
* 16135: gap of unknown length
* 16136
* 19100: contig of 2965 bp in length
* 19101
* 19200: gap of unknown length
* 19201
* 22184: contig of 2984 bp in length
* 22185
* 22284: gap of unknown length
* 22285
* 25287: contig of 3003 bp in length
* 25288
* 25387: gap of unknown length
* 25388
* 28431: contig of 3044 bp in length
* 28432
* 28531: gap of unknown length
* 28532
* 31409: contig of 2878 bp in length
* 31410
* 31509: gap of unknown length
* 31510
* 35320: contig of 3811 bp in length
* 35321
* 35420: gap of unknown length
* 35421
* 38808: contig of 3388 bp in length
* 38809
* 38908: gap of unknown length
* 43846: contig of 4938 bp in length
* 43847
* 43946: gap of unknown length
* 43947
* 49171: gap of unknown length
* 49172
* 53157: contig of 3886 bp in length
* 53158
* 53257: gap of unknown length
* 53258
* 57731: contig of 4474 bp in length
* 57732
* 57831: gap of unknown length
* 57832
* 62660: contig of 4829 bp in length
* 62661
* 62760: gap of unknown length
* 62761
* 68243: contig of 5483 bp in length
* 68244
* 68343: gap of unknown length
* 68344
* 73047: contig of 4704 bp in length
* 73048
* 73147: gap of unknown length
* 73148
* 78565: contig of 5418 bp in length
* 78566
* 78665: gap of unknown length
* 78666
* 85726: contig of 7061 bp in length
* 85727
* 85826: gap of unknown length
* 85827
* 91985: contig of 6159 bp in length
* 91986
* 92085: gap of unknown length
* 92086
* 97984: contig of 5899 bp in length
* 97985
* 98084: gap of unknown length
* 98085
* 105338: contig of 7254 bp in length
* 105339
* 105439: gap of unknown length
* 105440
* 113368: contig of 7930 bp in length
* 113369
* 113469: gap of unknown length
* 113470
* 120754: contig of 7285 bp in length
* 120755
* 120854: contig of 8489 bp in length
* 120855
* 129343: contig of 9556 bp in length
* 129344
* 138998: contig of 9556 bp in length
* 138999
* 139098: gap of unknown length
* 139099
* 149579: contig of 10481 bp in length
* 149580
* 149679: gap of unknown length

```

```

* 149580 162343: contig of 12664 bp in length
* 162344 162443: gap of unknown length
* 162444 185806: contig of 23363 bp in length.
FEATURES
  Location/Qualifiers
    1..185806
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /chromosome="UNK"
      /clone="RP23-459M2"
    1..1042
      /note="assembly_name:Contig24"
    1143..2216
      /note="assembly_name:Contig25"
    2317..3838
      /note="assembly_name:Contig26"
    3939..5396
      /note="assembly_name:Contig27"
    5497..6874
      /note="assembly_name:Contig28"
    6975..8481
      /note="assembly_name:Contig29"
    8582..9874
      /note="assembly_name:Contig30"
    9975..11483
      /note="assembly_name:Contig31"
    11584..13834
      /note="assembly_name:Contig32"
    13935..16035
      /note="assembly_name:Contig33"
    16136..19100
      /note="assembly_name:Contig34"
    19201..22184
      /note="assembly_name:Contig35"
    22285..25287
      /note="assembly_name:Contig36"
    25388..28431
      /note="assembly_name:Contig37"
    28532..31409
      /note="assembly_name:Contig38"
    31510..35320
      /note="assembly_name:Contig39"
    35421..38808
      /note="assembly_name:Contig40"
    38909..43846
      /note="assembly_name:Contig41"
    43947..49171
      /note="assembly_name:Contig42"
    49272..53157
      /note="assembly_name:Contig43"
    53258..57731
      /note="assembly_name:Contig44"
    57832..62660
      /note="assembly_name:Contig45"
    62761..68243
      /note="assembly_name:Contig46"
    68344..73047
      /note="assembly_name:Contig47"
    73148..78565
      /note="assembly_name:Contig48"
    78666..85726
      /note="assembly_name:Contig49"
    85827..91985
      /note="assembly_name:Contig50"
    92086..97984
      /note="assembly_name:Contig51"
    clone_end:SP6
    vector_side:left
    98085..105338
      /note="assembly_name:Contig52"
    105439..113368
      /note="assembly_name:Contig53"
    113469..120753
      /note="assembly_name:Contig54"

```

```

misc_feature 120854..129342
              /note="assembly_name:Contig55"
misc_feature 129443..138998
              /note="assembly_name:Contig56"
misc_feature 139099..149579
              /note="assembly_name:Contig57"
misc_feature 149680..162343
              /note="assembly_name:Contig58"
misc_feature 162444..185806

Query Match      6.3%; Score 92; DB 2; Length 185806;
Best Local Similarity 100.0%; Pred. No. 1.2e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGTGGATGGCTCGCGGTGTCTCGCCACCTTCGCGGATGAGCGCCAACTTACAAG 821
      |||||||
Db 159127 GCGTGGATGGCTCGCGGTGTCTCGCCACCTTCGCGGATGAGCGCCAACTTACAAG 159186

QY 822 ATCGAGACCTCGCTTCGCCACCAACTACAT 853
      |||||||
Db 159187 ATCGAGACCTCGCTTCGCCACCAACTACAT 159218

RESULT 9
AF303002      AF303002      6123 bp      DNA      linear      PRI 13-NOV-2001
LOCUS      Homo sapiens neurogenin 2 gene, partial cds.
DEFINITION
ACCESSION      AF303002
VERSION      AF303002.1 GI:11875763
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
TITLE      Neurogenin2 expression in ventral and dorsal spinal neural tube
progenitor cells is regulated by distinct enhancers
JOURNAL      Dev. Biol. 229 (2), 327-339 (2001)
MEDLINE      21077024
PUBMED      11203697
REFERENCE
AUTHORS      Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
TITLE      Direct Submission
JOURNAL      Submitted (06-SEP-2000) Center for Basic Neuroscience - NA4.146, UT
Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
75390-9111, USA
FEATURES
Source      Location/Qualifiers
1..6123
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>754
/product="neurogenin 2"
<1..754
/note="NGN2"
/codon_start=2
/product="neurogenin 2"
/protein_id="AAG40770.1"
/db_xref="GI:11875764"
/translation="ASPAALATPLSSADEEEEPGASGARRQGAEGQGGARGG
VAAGEGCRPARLLGVHDCKRRPSRAVSRGAKTAETVQRIKTRRLKANNRRNR
MHNLAALDALVELPTPEDAKLTETLRFAPHY IVALTETLLADHCGGGGLP
GALFSEVLLSPGASALSSGDSFSPASTWCTNSPAPSSVSNSTSPISCTLSP
ASPAQMDMYWPPPPDKHYAPHPFARDCI"
BASE COUNT      1484 a 1536 c 1507 g 1596 t
ORIGIN

Query Match      2.4%; Score 35; DB 9; Length 6123;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 853
      |||||||

```

Db 383 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 417

RESULT 10  
AC079846\_3/c  
WPCOMMENT

Sequence split into 4 fragments LOCUS AC079846 Accession AC079846

Fragment Name Begin End  
AC079846\_0 1 110000  
AC079846\_1 100001 210000  
AC079846\_2 200001 310000  
AC079846\_3 300001 391531

Continuation (4 of 4) of AC079846 from base 300001 (AC079846 Homo sapiens chromosome

Query Match 2.4%; Score 35; DB 2; Length 91531;

Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 853

|||||

Db 86995 AAGATCGAGACCTCGCTTCGCCACCAACTACAT 86961

RESULT 11  
AC023886/c

LOCUS AC023886 179697 bp DNA linear PRI 20-MAR-2002  
DEFINITION Homo sapiens BAC clone RP11-402J6 from 4, complete sequence.

ACCESSION AC023886

VERSION AC023886.7 GI:19482381

KEYWORDS HTG.

SOURCE Homo sapiens.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE Sulston,J.E. and Waterston,R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

JOURNAL 99063792

MEDLINE 9847074

PUBMED 2 (bases 1 to 179697)

REFERENCE Radionenko,M. and Abbott,A.

The sequence of Homo sapiens BAC clone RP11-402J6

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 179697)

AUTHORS Waterston,R.H.

TITLE Direct Submission

Submitted (18-FEB-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE 4 (bases 1 to 179697)

AUTHORS Waterston,R.H.

TITLE Direct Submission

Submitted (15-MAR-2002) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE 5 (bases 1 to 179697)

AUTHORS Waterston,R.

TITLE Direct Submission

Submitted (20-MAR-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Mar 15, 2002 this sequence version replaced gi:17352441.

COMMENT ----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: sapiens@wustl.edu

----- Summary Statistics

Center project name: H\_NH0402J06

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC004049, 2000 bp overlap; the clone sequenced to the right is RP11-148B6. Actual end of this clone is at base position 179697 of RP11-402J6.

#### FEATURES

```

source
1..179697
/db_xref="taxon:9606"
/map="4"
/clone="RP11-402J6"
/clone_lib="RPCI-11"
1..749
/rpt_family="L1"
734..3067
/rpt_family="L1"
3104..3444
/rpt_family="MaLR"
3447..4289
/rpt_family="L1"
4290..4369
/rpt_family="(TA)n"
4370..4797
/rpt_family="L1"
5022..5194
/rpt_family="L2"
5214..5628
/rpt_family="L1"
5710..5729
/rpt_family="(TA)n"
5749..5863
/rpt_family="(TA)n"
5883..6263
/rpt_family="ERVL"
9901..10151
/rpt_family="ERVL"
10152..10310
/rpt_family="L1"
10311..10400
/rpt_family="Mariner"
10401..10816
/rpt_family="L1"
10967..12362
/rpt_family="L1"
12366..12653
/rpt_family="Alu"

repeat_region 13695..13940 /rpt_family="L1"
repeat_region 13965..14272 /rpt_family="Alu"
repeat_region 14274..14795 /rpt_family="ERVL"
repeat_region 14806..15392 /rpt_family="L1"
repeat_region 21504..22112 /rpt_family="L1"
repeat_region 22131..23166 /rpt_family="L1"
repeat_region 23246..23386 /rpt_family="MaLR"
repeat_region 23511..23654 /rpt_family="MIR"
repeat_region 23655..25486 /rpt_family="L1"
repeat_region 25487..25636 /rpt_family="MIR"
repeat_region 25814..26124 /rpt_family="Alu"
repeat_region 26444..26592 /rpt_family="MIR"
repeat_region 27423..27713 /rpt_family="Alu"
repeat_region 27716..27739 /rpt_family="AT-rich"
repeat_region 27773..27795 /rpt_family="AT-rich"
repeat_region 28020..28429 /rpt_family="ERVL"
repeat_region 29646..29668 /rpt_family="AT-rich"
repeat_region 29727..29871 /rpt_family="L1"
repeat_region 29875..30114 /rpt_family="Alu"
repeat_region 30134..30267 /rpt_family="L1"
repeat_region 30268..30549 /rpt_family="L1"
repeat_region 30550..30883 /rpt_family="MaLR"
repeat_region 32461..32791 /rpt_family="MaLR"
repeat_region 32792..33487 /rpt_family="L1"
repeat_region 33488..33847 /rpt_family="MaLR"
repeat_region 33848..34354 /rpt_family="L1"
repeat_region 34358..34379 /rpt_family="(CATATA)n"
repeat_region 34381..34502 /rpt_family="Alu"
repeat_region 34503..34556 /rpt_family="L1"
repeat_region 34557..34859 /rpt_family="Alu"
repeat_region 34860..35041 /rpt_family="L1"
repeat_region 35219..35676 /rpt_family="L2"
repeat_region 35974..36042 /rpt_family="L2"
repeat_region 36833..37144 /rpt_family="Alu"
repeat_region 36833..36848 /note="match to EST AI537116 (NTD:g4451251) to15h09.x1"
repeat_region 37147..37410 /rpt_family="Alu"
repeat_region 38154..38424 /rpt_family="Alu"

```

```

repeat_region      /rpt_family="Alu"
                    39941..40230
misc_feature        /rpt_family="Alu"
                    39941..39956
                    /note="match to EST AA775240 (NID:92834574) ad18a05.s1"
repeat_region      40767..40798

Query Match
Best Local Similarity 2.4%; Score 35; DB 9; Length 179697;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTGGCTTCGCCACACACTACAT 853
      |||||||
Db 55815 AAGATCGAGACCTGGCTTCGCCACACACTACAT 55781

RESULT 12
LOCUS      HSA133776      1330 bp      DNA      linear      PRI 19-JUN-1999
DEFINITION Homo sapiens gene for neurogenin 3.
ACCESSION  AJ133776
VERSION     AJ133776.1 GI:5123782
KEYWORDS   bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1330)
AUTHORS     Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and
            Mallet,J.
TITLE       The human neurogenin 3 homolog maps to chromosome 10q21.3 and its
            expression pattern is identical to that of its murine counterparts
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1330)
AUTHORS     Ravassard,P.
TITLE       Direct Submission
JOURNAL     Submitted (16-MAR-1999) Ravassard P., Lpn, CNRS UMRC 9923, Hopital
            de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013
            PARIS, FRANCE

FEATURES             Location/Qualifiers
     source           1..1330
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene             1..1330
                     /gene="ngn3"
     mRNA             join(<1..157,321..>1330)
                     /gene="ngn3"
     exon             1..157
                     /gene="ngn3"
                     /number=1
     intron           158..320
                     /gene="ngn3"
                     /number=1
     exon             321..1330
                     /gene="ngn3"
                     /number=2
     CDS              322..966
                     /gene="ngn3"
                     /function="early neurogenesis"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="CA845384.1"
                     /db_xref="GI:5123783"
                     /translation="MTPOPSGAPTVQVTRERSFPRASEDEVCTSPPTPTPG
            NCAAEEGCGAPRKLRARGSRPKSELALSKORRRKKANDRRNRHMLNSA
            LDALRGVLPFPDDAKLTETLRFAHNYIWAQTQIRADHSLYALEPPAPHCGLG
            SPGGSPGWSGLYSPVQAGSLSPAASLERPGLLGATSSACLSGSLAFSDFL"

BASE COUNT      230 a 459 c 413 g 228 t
ORIGIN
Query Match
Best Local Similarity 2.2%; Score 32; DB 9; Length 1330;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 672 AAGAGCGAGTTGGCACTGAGCAGCAGCGACG 703
      |||||||
Db 535 AAGAGCGAGTTGGCACTGAGCAGCAGCGACG 566

RESULT 13
LOCUS      AF234829      5340 bp      DNA      linear      PRI 19-OCT-2001
DEFINITION Homo sapiens neurogenin 3 gene, complete cds.
ACCESSION  AF234829
VERSION     AF234829.1 GI:13183002
KEYWORDS
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 5340)
AUTHORS     del Bosque-Plata,L., Lin,J., Horikawa,Y., Schwarz,P.E., Cox,N.J.,
            Iwasaki,N., Ogata,M., Iwamoto,Y., German,M.S. and Bell,G.I.
TITLE       Mutations in the coding region of the neurogenin 3 gene (NEUROG3)
            are not a common cause of maturity-onset diabetes of the young in
            Japanese subjects
JOURNAL     Diabetes 50 (3), 694-696 (2001)
MEDLINE     21140923
PUBMED      11246894
REFERENCE   2 (bases 1 to 5340)
AUTHORS     Lin,J. and German,M.
TITLE       Direct Submission
JOURNAL     Submitted (15-FEB-2000) Hormone Research Institute, University of
            California San Francisco, 513 Parnassus Ave., San Francisco, CA
            94143-0534, USA

FEATURES             Location/Qualifiers
     source           1..5340
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     mRNA            <3022..>3666
                     /product="neurogenin 3"
                     /codon_start=1
     CDS              3022..3666
                     /product="neurogenin 3"
                     /protein_id="AAK15022.1"
                     /db_xref="GI:13183003"
                     /translation="MTPOPSGAPTVQVTRERSFPRASEDEVCTSPPTPTPTRG
            NCAAEEGCGAPRKLRARGSRPKSELALSKORRRKKANDRRNRHMLNSA
            LDALRGVLPFPDDAKLTETLRFAHNYIWAQTQIRADHSLYALEPPAPHCGLG
            SPGGSPGWSGLYSPVQAGSLSPAASLERPGLLGATSSACLSGSLAFSDFL"

BASE COUNT      1215 a 1500 c 1514 g 1111 t
ORIGIN
Query Match
Best Local Similarity 2.2%; Score 32; DB 9; Length 5340;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCACTGAGCAGCAGCGACG 703
      |||||||
Db 3235 AAGAGCGAGTTGGCACTGAGCAGCAGCGACG 3266

RESULT 14
LOCUS      AL450311/c      165110 bp      DNA      linear      PRI 12-JUL-2001
DEFINITION Human DNA sequence from clone RP11-34J33 on chromosome 10, complete
            sequence.
ACCESSION  AL450311
VERSION     AL450311.11 GI:14626972
KEYWORDS   HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 165110)
AUTHORS     Howden,P.

```



TITLE  
JOURNAL

## Direct Submission

Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humqueres@sanger.ac.uk

## COMMENT

On Jul 8, 2001 this sequence version replaced gi:14575291.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit subsequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr10>  
RP11-34333 is from the library RP11-2 constructed by the group  
of Pieter de Jong. For further details see

<http://www.chori.org/dacpac/home.htm>

VECTOR: pBACe3.6

## FEATURES

## source

1. .165110

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="10"

/clone="RP11-34333"

/clone\_lib="RP11-2"

7. .147

/notes="THE1C repeat: matches 2. .142 of consensus"

151. .293

/notes="MIR repeat: matches 2. .148 of consensus"

1136. .1187

/note="26 copies 2 mer gt 98% conserved"

2960. .3272

/notes="AluSg repeat: matches 1. .310 of consensus"

4144. .4270

/notes="AluX repeat: matches 1. .134 of consensus"

4319. .4494

/note="AluX repeat: matches 118. .293 of consensus"

5862. .6438

/notes="L1MEC repeat: matches 272. .1095 of consensus"

6916. .8080

/notes="L1MEC repeat: matches 1168. .2367 of consensus"

9934. .10087

/note="MIR repeat: matches 25. .184 of consensus"

10353. .10481

/notes="MIR repeat: matches 65. .194 of consensus"

11077. .11372

/notes="AluX repeat: matches 1. .294 of consensus"

11600. .11639

/note="20 copies 2 mer tc 95% conserved"

12479. .12608

/note="65 copies 2 mer at 90% conserved"

13433. .13876

/notes="MLT1C repeat: matches 1. .466 of consensus"

14372. .14576

/note="AluSg repeat: matches 1. .301 of consensus"

14577. .14633

/note="AluSg repeat: matches 243. .299 of consensus"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

14843. .14943

/note="MER86 repeat: matches 9. .109 of consensus"

15404. .15613

/note="MIR repeat: matches 3. .213 of consensus"

16626. .16705

/note="MIR repeat: matches 60. .140 of consensus"

16928. .16993

/note="33 copies 2 mer gg 66% conserved"

17476. .17569

/note="MER81 repeat: matches 2. .114 of consensus"

17719. .18069

/note="L1MC4 repeat: matches 7617. .7977 of consensus"

18117. .18345

/note="MIR repeat: matches 7. .262 of consensus"

19169. .19400

/note="MIR repeat: matches 5. .239 of consensus"

19436. .19627

/note="L1MC5 repeat: matches 7728. .7917 of consensus"

19642. .19716

/note="MIR repeat: matches 48. .131 of consensus"

19699. .20010

/note="21 copies 2 mer tg 100% conserved"

20623. .20704

/note="41 copies 2 mer gt 85% conserved"

20738. .20773

/note="U2 repeat: matches 1. .36 of consensus"

20802. .20863

/note="MLT1J repeat: matches 1. .62 of consensus"

20889. .21263

/note="MLT1F repeat: matches 188. .541 of consensus"

21463. .21618

/note="MIR repeat: matches 46. .192 of consensus"

22019. .22326

/note="AluB repeat: matches 1. .306 of consensus"

22381. .22564

/note="L2 repeat: matches 2453. .2629 of consensus"

22896. .23174

/note="MLT1J repeat: matches 117. .413 of consensus"

23215. .23346

/note="MIR repeat: matches 48. .188 of consensus"

23388. .23532

/note="L2 repeat: matches 2097. .2230 of consensus"

23533. .23837

/note="AluB repeat: matches 1. .303 of consensus"

23838. .24137

/note="L2 repeat: matches 1754. .2097 of consensus"

24291. .24581

/note="AluX repeat: matches 1. .300 of consensus"

24653. .24850

/note="MIR repeat: matches 1. .200 of consensus"

26189. .26685

/note="Cpg island"

/evidence-not\_experimental

29453. .30918

/note="Cpg island"

/evidence-not\_experimental

31084. .31156

/note="MIR repeat: matches 65. .138 of consensus"

31618. .31876

/note="AluJo repeat: matches 29. .275 of consensus"

32767. .32830

/note="MIR repeat: matches 76. .139 of consensus"

33050. .33178

/note="43 copies 3 mer tcc 72% conserved"

35112. .36201

/note="Cpg island"

/evidence-not\_experimental

36551. .36604

/note="27 copies 2 mer ac 94% conserved"

37585. .38254

/note="L1PA15 repeat: matches 5480. .6157 of consensus"

38265. .38561

/note="AluX repeat: matches 1. .300 of consensus"

```

repeat_region 39214..39423
/Note="MIR repeat: matches 3..219 of consensus"
repeat_region 39489..39801
/Note="AluSx repeat: matches 1..292 of consensus"
repeat_region 40050..40189
/Note="MIR repeat: matches 1..144 of consensus"
repeat_region 41057..41290
/Note="MIR repeat: matches 8..240 of consensus"
repeat_region 41296..41460
/Note="L2 repeat: matches 2569..2730 of consensus"
repeat_region 41506..41944
/Note="L2 repeat: matches 1916..2416 of consensus"
repeat_region 42388..42698
/Note="L2 repeat: matches 1448..1779 of consensus"
repeat_region 44193..44579
/Note="THEIC repeat: matches 1..371 of consensus"
repeat_region 44600..44956
/Note="LTR16A repeat: matches 90..445 of consensus"
repeat_region 45240..45300
/Note="MR58A repeat: matches 37..97 of consensus"
repeat_region 45798..45909
/Note="L2 repeat: matches 90..192 of consensus"
repeat_region 46826..46871
/Note="L2 repeat: matches 90..192 of consensus"
repeat_region 50684..50969
/Note="AluSg repeat: matches 1..287 of consensus"
repeat_region 50980..51291
/Note="AluSx repeat: matches 1..312 of consensus"
repeat_region 52222..52519
/Note="AluSx repeat: matches 1..300 of consensus"
repeat_region 54065..54260
/Note="L1M4 repeat: matches 3865..4055 of consensus"
repeat_region 54261..54432
/Note="FAM repeat: matches 2..167 of consensus"
repeat_region 54433..54629
/Note="L1M4 repeat: matches 3652..3865 of consensus"
repeat_region 54648..54862
/Note="L1M4 repeat: matches 11..217 of consensus"
repeat_region 54863..55236
/Note="L1M1 repeat: matches 1..365 of consensus"
repeat_region 55237..55700
/Note="L1M4 repeat: matches 217..716 of consensus"

Query Match 2.2%; Score 32; DB 9; Length 165110;
Best Local Similarity 100.0%; Pred. No. 8.6e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGAGCGAGTTGGCACTGAGCAGCGACG 703
|||||
Db 30112 AAGAGCGAGTTGGCACTGAGCAGCGACG 30081

RESULT 15
AC021954/c
LOCUS AC021954.3 173341 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT
SEQUENCE 24 unordered pieces.
AC021954
AC021954
AC021954.3 GI:7417809
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173341)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Triglilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5931
Center clone name: 57_E.12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 171041; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

TITLE  
JOURNALREFERENCE  
AUTHORSTITLE  
JOURNAL  
COMMENT

```

DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Triglilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 173341)
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Triglilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5931
Center clone name: 57_E.12
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 171041; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 11:26:20 ; Search time 374 Seconds

(without alignments)

8791.223 Million cell updates/sec

Title: US-09-595-947C-1

Perfect score: 1460

Sequence: 1 gcaggtagcgagagagcag.....agagtacctaataccagtgt 1460

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 8

Total number of hits satisfying chosen parameters: 2526306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1460	100.0	1491 19	RAV42512
2	92	6.3	804 19	AAV27050
3	92	6.3	804 21	AAZ51981
4	92	6.3	861 22	AAF27266
5	92	6.3	1861 21	AAC61090
6	92	6.3	5567 22	AAF27254
7	65	4.5	65 24	ABN31392
8	35	2.4	65 24	ABN57521
9	32	2.2	5340 21	AAC61089

c	10	29	2.0	428	22	AA333797	Human cDNA encoding Human Neurogenin-3
	11	26	1.8	26	24	ABT03700	PCR primer used to
	12	25	1.7	25	19	AAV42515	Mouse neurogenin 1
	13	23	1.6	738	19	AAV27046	Murine neurogenin-1
	14	23	1.6	738	21	AAZ51977	Chicken atonal hom
	15	23	1.6	790	22	AAF27264	DNA encoding murin
	16	23	1.6	1332	19	AAV42938	Mouse neurogenic d
	17	23	1.6	1333	18	AAT74894	Mouse neurogenin 2
	18	23	1.6	1385	19	AAV27049	Murine neurogenin-
	19	23	1.6	1385	21	AAZ51980	Mouse neurogenin 2
	20	23	1.6	1385	22	AAF27269	Mouse atonal homol
	21	23	1.6	1412	22	AAF27255	Mouse atonal homol
	22	23	1.6	1412	22	AAF27273	Human neurogenic d
	23	21	1.4	1268	18	AAT74891	DNA encoding human
	24	21	1.4	1268	19	AAV42932	Human neurogenic d
	25	21	1.4	1535	18	AAT74890	DNA encoding human
	26	21	1.4	1535	19	AAV42931	Human Neurod2 gene
	27	21	1.4	1550	21	AA62681	Human reproductive
	28	21	1.4	2776	22	AAAL04043	Human reproductive
	29	21	1.4	2776	22	AAAL04045	Human immune/haema
	30	21	1.4	2776	22	AAK68475	Human immune/haema
	31	21	1.4	2776	22	AAK68476	Oligonucleotide fo
	32	20	1.4	592	24	ABQ49522	Oligonucleotide fo
	33	20	1.4	592	24	ABQ49523	Acetylcholine tran
	34	19	1.3	50	17	AAT13245	NACR beta2 subuni
	35	19	1.3	50	17	AAT28933	M. capsulatus gene
	36	19	1.3	510	24	ABO91538	M. capsulatus gene
	37	19	1.3	615	24	ABO91539	M. capsulatus gene
	38	19	1.3	1260	24	ABQ91700	Arabidopsis thalia
	39	19	1.3	1263	21	AAQ3008	Arabidopsis thalia
	40	19	1.3	1698	21	AAC36256	Human reproductive
	41	19	1.3	3424	22	AAAL06199	Human testicular a
	42	19	1.3	3424	23	ABL98764	Human reproductive
	43	19	1.3	3425	22	AAAL06200	Human testicular a
	44	19	1.3	3425	23	ABL98765	Human immune/haema
	45	19	1.3	18733	22	AAK80682	Human immune/haema

#### ALIGNMENTS

RESULT 1  
AAV42512  
ID AAV42512 standard; cDNA; 1491 BP.  
XX  
AC AAV42512;  
XX  
DT 05-OCT-1998 (first entry)  
XX  
DE cDNA encoding a novel BHLH protein designated RELAX.  
XX  
KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;  
KW control; gene expression; transcriptional activator; targeting;  
KW protein expression; central nervous system; CNS; treatment;  
KW nervous system disorder; ss.  
XX  
OS Rattus sp.  
XX  
FH Key  
FT CDS  
FT Location/Qualifiers  
FT 459..1103  
FT /\*tag= a  
FT /product= RELAX  
XX  
PN WO9827206-A2.  
XX  
PD 25-JUN-1998.  
XX  
PF 19-DEC-1997; 97WO-FR02368.  
XX  
PR 19-DEC-1996; 96FR-0015651.  
XX  
PA (RHON ) RHONE-POULENC RORER SA.  
XX

PI Mallet J, Ravassard P, Icard-Liepkalns C;

XX WPI; 1998-362775/31.  
DR P-PSDB; AAW62991.

XX Basic helix-loop-helix polypeptide and related nucleic acid - with  
PT transcriptional activity, for targeting expression of genes to  
PT central nervous system and treatment of nervous disease

XX PS Claim 6; Page 20; 28pp; French.

XX CC The present sequence encodes a basic helix-loop-helix (BHLH) type  
CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.  
CC The protein is used to control and participate in gene expression,  
CC by acting as a transcriptional activator, strictly dependent on the  
CC presence of an intact E box (CANNTG), particularly for targeting  
CC expression of proteins to the central nervous system (CNS). The  
CC nucleic acid sequence can be used to treat nervous system disorders,  
CC and antisense sequences can be used to control mRNA transcription.

XX SQ Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;

Query Match 100.0%; Score 1460; DB 19; Length 1491;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCGCCCTGCTGATTTGGCCGCTGGCAGGCA 60  
DB 1 GCAGGTAGCAGAGGAGCAGTCCCTGGGCCCGCCCTGCTGATTTGGCCGCTGGCAGGCA 60  
QY 61 GCAGCCCGCAGGAGCAGTCCCTGGTCCGGGAGAGCAGATTAAGCGTGGCAGGAGCACA 120  
DB 61 GCAGCCCGCAGGAGCAGTCCCTGGTCCGGGAGAGCAGATTAAGCGTGGCAGGAGCACA 120  
QY 121 CGATTAGCAGTCAAGTCCCTGGTCTCAGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 180  
DB 121 CGATTAGCAGTCAAGTCCCTGGTCTCAGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 180  
QY 181 CCGAGCTCTTTGCTGCTCCAGAGCAGCAATTTACTCAGGCGAGGCGCCCTGCAGCTCAG 240  
DB 181 CCGAGCTCTTTGCTGCTCCAGAGCAGCAATTTACTCAGGCGAGGCGCCCTGCAGCTCAG 240  
QY 241 CAAAGTTCGAGGAGCAGAGGGGTTTCAAGTATCCACCGTGTGCTGCTGCTGCTGCTGCTG 300  
DB 241 CAAAGTTCGAGGAGCAGAGGGGTTTCAAGTATCCACCGTGTGCTGCTGCTGCTGCTGCTG 300  
QY 301 GCAGCTCTCTCTTTGAGCCGAGTAACTAGGTAACTAGGTAACTAGGTAACTAGGTAACTAG 360  
DB 301 GCAGCTCTCTCTTTGAGCCGAGTAACTAGGTAACTAGGTAACTAGGTAACTAGGTAACTAG 360  
QY 361 TAGAAGAGGGAGTGGTGGCGTACTAGTCCCGCGTGGAGTGAACCTCTAAGTCAAG 420  
DB 361 TAGAAGAGGGAGTGGTGGCGTACTAGTCCCGCGTGGAGTGAACCTCTAAGTCAAG 420  
QY 421 ACTGTACACCCCTCCATTTTCCCAACCTCAGGATGGCGCTCATCCCTTGGATG 480  
DB 421 ACTGTACACCCCTCCATTTTCCCAACCTCAGGATGGCGCTCATCCCTTGGATG 480  
QY 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGCCTCGGACACG 540  
DB 481 CGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGGAGCCTCGGACACG 540  
QY 541 AAGTGTCTCAGTTCCTCAATTTCCACCCACTAGTCCCTCTCTCTCTCTCTCTCTCTCTCT 600  
DB 541 AAGTGTCTCAGTTCCTCAATTTCCACCCACTAGTCCCTCTCTCTCTCTCTCTCTCTCTCT 600  
QY 601 AAGCAGAGCAGTGTACTGCGGAGGAGCATCGAGGAAGTCTGCTGCGGCGCGGAGGGC 660  
DB 601 AAGCAGAGCAGTGTACTGCGGAGGAGCATCGAGGAAGTCTGCTGCGGCGCGGAGGGC 660  
QY 661 GCAACAGCCCAAGAGCGAGTGGTCACTGAGCAGCAGCAGGAGCGCGCGCAGGAAGG 720  
DB 661 GCAACAGCCCAAGAGCGAGTGGTCACTGAGCAGCAGCAGGAGCGCGCGCAGGAAGG 720

QY 721 CCAACGACCGGGAGCGCAACCGCATGACACACCTTAACCTCCGCGCTGGATCGCTGGCGG 780  
DB 721 CCAACGACCGGGAGCGCAACCGCATGACACACCTTAACCTCCGCGCTGGATCGCTGGCGG 780  
QY 781 GTGTCTCTCCACCTTCCCGGATGACGCCAACTTACAAAGATCGAGACCTCGCTGGCTTCG 840  
DB 781 GTGTCTCTCCACCTTCCCGGATGACGCCAACTTACAAAGATCGAGACCTCGCTGGCTTCG 840  
QY 841 CCCAACTACATTTGGGCACTGACTCAGACGCTGGCGATAGCGGACACACAGCTTCTACG 900  
DB 841 CCCAACTACATTTGGGCACTGACTCAGACGCTGGCGATAGCGGACACACAGCTTCTACG 900  
QY 901 GCCCGGACCGCCCTGTCCTCTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGCG 960  
DB 901 GCCCGGACCGCCCTGTCCTCTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGCG 960  
QY 961 ACTGGGCTCTATCTACTCTCCAGTTCCTCCAAAGTGGTAGCTGAGCCCCACAGCTCAT 1020  
DB 961 ACTGGGCTCTATCTACTCTCCAGTTCCTCCAAAGTGGTAGCTGAGCCCCACAGCTCAT 1020  
QY 1021 TGGAGGAGTTCCTGGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1080  
DB 1021 TGGAGGAGTTCCTGGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1080  
QY 1081 TGGTGTCTCAGACTTCTTGTGAAGGGCCCAACAGGCGCTGGGCTGGGCTGGGCG 1140  
DB 1081 TGGTGTCTCAGACTTCTTGTGAAGGGCCCAACAGGCGCTGGGCTGGGCTGGGCG 1140  
QY 1141 AAAGGGAGGAGTCTGCTGCTGAATGAAGTAGTGGAGGAGTCTGAGGAGTCTGCTGCTG 1200  
DB 1141 AAAGGGAGGAGTCTGCTGCTGAATGAAGTAGTGGAGGAGTCTGAGGAGTCTGCTGCTG 1200  
QY 1201 CCTTCTGGCTTCTATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
DB 1201 CCTTCTGGCTTCTATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
QY 1261 GCTGTGGCTGACAAAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
DB 1261 GCTGTGGCTGACAAAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320  
QY 1321 ACCTCAACTCCGCTCCCAAGCAGAGGAGCGCTAGCAGTAAATAGTTGGAGAGTCTCC 1380  
DB 1321 ACCTCAACTCCGCTCCCAAGCAGAGGAGCGCTAGCAGTAAATAGTTGGAGAGTCTCC 1380  
QY 1381 ATACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
DB 1381 ATACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440  
QY 1441 AGAGTGAGCTTAATCCAGTGT 1460  
DB 1441 AGAGTGAGCTTAATCCAGTGT 1460

## RESULT 2

AAV27050  
ID AAV27050 standard; cDNA; 804 BP.

XX AAV27050;

XX 17-SEP-1998 (first entry)

XX Mouse neurogenin 3 gene.

ds; Mouse; neurogenin; expression vector; recombinant protein;  
antibody; neurogenesis.

XX Mus sp.

XX Key Location/Qualifiers  
FT CDS 160..804  
FT /\*tag= a  
FT /product= "Mouse neurogenin 3"





XX 01-JUN-2000; 2000WO-US15410.  
XX  
XX  
PR 01-JUN-1999; 99US-0137060.  
PR 19-JAN-2000; 2000US-0176993.  
XX  
XX  
PA (BAYU ) BAYLOR COLLEGE MEDICINE.  
XX  
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
XX  
XX  
DR WPI; 2001-032190/04.  
DR P-PSDB; AAB60350.  
XX  
XX  
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or  
PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
PT osteoarthritis and abnormal cell proliferation -  
XX  
XX  
PS Disclosure; Page -; 142pp; English.  
XX  
XX  
CC The invention relates to the use of atonal-associated nucleic acid or  
CC amino acid sequence, or any of its homologues or orthologues as  
CC therapeutic agents for the treatment of deafness, partial hearing loss,  
CC vestibular effects due to damage or loss of inner hair cells,  
CC osteoarthritis and abnormal cell proliferation. The invention also  
CC encompasses methods of screening for compounds which affect the  
CC expression of an atonal-associated nucleic acid sequence in an animal,  
CC and a transgenic animal in which an allele of a native atonal-associated  
CC gene is replaced by a heterologous nucleic acid sequence, thus  
CC inactivating the atonal-associated allele. The nucleic acids or proteins  
CC may be used in a method of treating an animal for hearing impairment,  
CC joint disease, balance disorders, abnormal cell proliferation, or other  
CC disease related to loss of a functional atonal-associated nucleic acid or  
CC protein. They may particularly be used to treat an animal with a  
CC deficiency in cerebellar granule neurons or their precursors, and may  
CC also be used in promoting mechanoreceptive cell growth and generating  
CC hair cells. The present sequence represents an atonal-associated nucleic  
CC acid sequence referred to in the invention.  
CC Note: The present sequence is not shown in the specification, but  
CC was obtained from Genbank.  
XX  
SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;  
Query Match 6.3%; Score 92; DB 22; Length 5567;  
Best Local Similarity 100.0%; Pred. No. 8.5e-35;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 762 GCCTGGATGCGTGGCGGTGCTGCTGCCACCTTCCCGGATGACGCCAACTTACAAG 821  
DB 5226 GCCTGGATGCGTGGCGGTGCTGCTGCCACCTTCCCGGATGACGCCAACTTACAAG 5285  
QY 822 ATCGAGACCTGGCTGCGCCACAACTACAT 853  
DB 5286 ATCGAGACCTGGCTGCGCCACAACTACAT 5317  
RESULT 7  
ABN31392  
ID ABN31392 standard; DNA; 65 BP.  
XX  
XX  
AC ABN31392;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Rat spliced transcript detection oligonucleotide SEQ ID NO:4140.  
XX  
KW Human; mouse; rat; splice transcript; detection; RNA transcript;  
KW splice variant; transcriptome; oligonucleotide library; ss.  
XX  
XX Rattus norvegicus.  
OS  
XX WO200210449-A2.  
XX  
XX 07-FEB-2002.  
PD

XX 20-JUL-2001; 2001WO-IB01903.  
XX  
XX  
PR 28-JUL-2000; 2000US-221607P.  
PR 02-MAY-2001; 2001US-287724P.  
XX  
XX  
PA (COMP-) COMPUGEN INC.  
XX  
XX  
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
XX  
XX  
DR WPI; 2002-257383/30.  
XX  
XX  
PT New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes -  
XX  
XX  
PS Example 1; SEQ ID 4140; 47pp; English.  
XX  
XX  
CC The present invention describes oligonucleotide libraries for detecting  
CC messenger RNAs that populate a (sub-)transcriptome, where the  
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
CC transcription units that populate a genome. The library comprises  
CC several oligonucleotides, each capable of hybridising selectively to a  
CC set of messenger RNAs transcribed from a given transcription unit of  
CC the genome, which encodes one or more messenger RNA splice variants.  
CC The oligonucleotide libraries are useful for detecting mRNAs from a  
CC biological sample, in expression profiling studies, in qualitatively or  
CC quantitatively characterising the corresponding transcriptome, and in  
CC detecting RNA transcripts and splice variants of human or animal  
CC transcriptomes. The libraries may also be used as specialised mini  
CC libraries to detect transcripts of a sub-transcriptome under a  
CC particular biological or pathological state, and so allowing the  
CC detection of tissue- and pathology-specific genes such as those genes  
CC only expressed in specific tissue under a specific pathological  
CC condition; to detect developmental specific genes; and to detect RNA  
CC transcripts and splice variants of a transcriptome of a patient suffering  
CC from a particular disorder. ABN27253 to ABN59589 represent  
CC oligonucleotide sequences from rats, humans and mice, which are used in  
CC the exemplification of the present invention.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 65 BP; 11 A; 20 C; 14 G; 20 T; 0 other;  
Query Match 4.5%; Score 65; DB 24; Length 65;  
Best Local Similarity 100.0%; Pred. No. 1.8e-21;  
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1252 TTCCTTGCTGCTGCTGCACAAAGGACATTCGAGCTGATCCTTAAACCTCCTC 1311  
DB 1 TTCCTTGCTGCTGCTGCACAAAGGACATTCGAGCTGATCCTTAAACCTCCTC 60  
QY 1312 AGTGT 1316  
DB 61 AGTGT 65  
RESULT 8  
ABN57521  
ID ABN57521 standard; DNA; 65 BP.  
XX  
XX  
AC ABN57521;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:30269.  
XX  
KW Human; mouse; rat; splice transcript; detection; RNA transcript;  
KW splice variant; transcriptome; oligonucleotide library; ss.  
XX  
XX Mus musculus.  
OS



```

XX PN WO200210449-A2.
XX XX
XX PD 07-FEB-2002.
XX XX
XX PF 20-JUL-2001; 2001WO-IB01903.
XX XX
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX XX
XX FA (COMP-) COMPUGEN INC.
XX XX
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX XX
XX DR WPI; 2002-257383/30.
XX XX
XX PT New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes.
XX XX
XX PS Example 1; SEQ ID 30269; 47pp; English.
XX CC
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridising selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pat_sequences.
XX XX
XX SQ Sequence 65 BP; 16 A; 25 C; 11 G; 13 T; 0 other;

Query Match 2.4%; Score 35; DB 24; Length 65;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 547 TCAGTTCCTCAATCCACCCACCTAGCCCTCTC 581
Db 1 TCAGTTCCTCAATCCACCCACCTAGCCCTCTC 35
|||||

RESULT 9
AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX XX
XX AC AAC61089;
XX XX
XX DT 05-FEB-2001 (first entry)
XX XX
XX DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX XX
XX KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
XX KW islet cell precursor identification; diabetes mellitus; human; ds.
XX OS Homo sapiens.

```

```

XX FH Key Location/Qualifiers
XX FT CDS 3022..3666
XX FT /*tag= a
XX FT /product= "Ngn3"
XX FT /note= "Neurogenin 3"
XX PN WO200059936-A1.
XX XX
XX PD 12-OCT-2000.
XX XX
XX PF 28-MAR-2000; 2000WO-US08436.
XX XX
XX PR 06-APR-1999; 99US-0128180.
XX XX
XX FA (REGC ) UNIV CALIFORNIA.
XX XX
XX PI German MS, Lin J;
XX XX
XX DR WPI; 2000-664989/64.
XX DR P-PSDB; AAY85617.
XX XX
XX PT Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX PT them, useful for diagnosis, prevention and treatment of diabetes
XX PT mellitus and to identify individuals at risk of diabetes.
XX PS Claim 6; Page 46-48; 54pp; English.
XX XX
XX CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX CC protein AAY85617. The Ngn3 gene is located at chromosome position
XX CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX CC protein sequences, and includes an antibody recognising the Ngn3 protein.
XX CC Also included in the invention is a method for identifying an islet cell
XX CC precursor, the method involves analysing a cell for the expression of the
XX CC Ngn3 gene product, where detection of the product is indicative of an
XX CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX CC reagent for detecting (in a subject) a predisposition to a defect in
XX CC pancreatic islet cell function or formation associated with a defect in
XX CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX CC precursor cells expressing Ngn3, and to alter cellular differentiation in
XX CC culture in vivo to produce new beta-cells to treat patients with diabetes
XX CC mellitus.
XX XX
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Query Match 2.2%; Score 32; DB 21; Length 5340;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 AAGACGCGAGTTGGCACTGACGAAGCAGCGACG 703
Db 3235 AAGACGCGAGTTGGCACTGACGAAGCAGCGACG 3266
|||||

RESULT 10
AAS33797
ID AAS33797 standard; cDNA; 428 BP.
XX XX
XX AC AAS33797;
XX XX
XX DT 17-DEC-2001 (first entry)
XX XX
XX DE Human cDNA encoding a novel foetal antigen, SEQ ID No 321.
XX XX
XX KW Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
XX KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
XX KW cardiovascular; autoimmune disease; rheumatoid arthritis;
XX KW hyperproliferative disorder; breast neoplasm; cancer;
XX KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX KW cerebral ischaemia; angiogenesis; nervous system disorder;
XX KW Alzheimer's disease; infection; ocular disorder; corneal infection;
XX KW wound healing; epithelial cell proliferation; food additive.
XX XX

```

OS Homo sapiens.  
XX WO200155312-A2.  
PN 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01321.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0198874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0228279.  
PR 22-AUG-2000; 2000US-0228681.  
PR 22-AUG-2000; 2000US-0228688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PA

XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-488782/53.  
 DR P-PSDB; AAU20977.  
 XX New polynucleotides and polypeptides for diagnosing, treating,  
 PT preventing or prognosing e.g. diseases or disorders of the nervous,  
 PT musculoskeletal, excretory, gastrointestinal, reproductive, and  
 PT respiratory systems  
 XX Claim 1; SEQ ID NO 321; 642pp; English.  
 XX The invention relates to novel nucleic acids encoding novel human foetal  
 CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
 CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. The antibodies to the antigens can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. Numerous  
 CC examples of diseases and disorders treated by the nucleic acids and  
 CC proteins are given in the specification. The present sequence

Query Match 2.0%; Score 29; DB 22; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 0.0007;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 825 GAGACCTGGCTGCGCCCACTACAT 853  
 |||||  
 Db 1 GAGACCTGGCTGCGCCCACTACAT 29

RESULT 11  
 ABT03700/C  
 ID ABT03700 standard; DNA; 26 BP.  
 AC ABT03700;  
 XX  
 XX 13-SEP-2002 (first entry)  
 DT  
 XX Human Neurogenin-3 gene PCR primer SEQ ID NO: 221.  
 DE

XX Human; cancer; neoplastic disease; tumour specific marker; cytostatic;  
 KW transcription factor; PCR; primer; ss.  
 KW  
 XX Homo sapiens.  
 OS

XX W0200240716-A2.  
 PN  
 XX 23-MAY-2002.  
 PD  
 XX 13-NOV-2001; 2001WO-US43461.  
 PF

XX 16-NOV-2000; 2000US-249508P.  
 PR  
 XX (CEMI-) CEMINES LLC.  
 PA  
 XX Palm K;  
 PI

XX WPI; 2002-537346/57.  
 XX  
 XX Determining the presence of neoplastic molecular markers, by  
 PT identifying the presence of markers in host test sample using array of  
 PT neoplastic molecular marker specific reagents and analyzing the array  
 of the reagents  
 XX Example 1; Page 17; 41pp; English.  
 XX  
 XX The present invention relates to a method for determining the presence of  
 CC neoplastic molecular markers in a host, involving the use of neoplastic  
 CC molecular marker specific reagents to detect such markers and analysing  
 CC the array of reagents, allowing the identification of the neoplastic  
 CC disease present. This can be used to determine the best treatment for  
 CC cancers, in particular neural cell, lung and prostate tumours. The  
 CC present sequence is a PCR primer useful for detecting the coding  
 CC sequences of markers of the invention.  
 XX  
 XX Sequence 26 BP; 3 A; 10 C; 5 G; 8 T; 0 other;  
 SQ  
 Query Match 1.8%; Score 26; DB 24; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 0.024;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 673 AGAGCGAGTTGGCACTGAGCAAGCAG 698  
 |||||  
 Db 26 AGAGCGAGTTGGCACTGAGCAAGCAG 1  
 RESULT 12  
 AAV42515  
 ID AAV42515 standard; DNA; 25 BP.  
 XX  
 XX AAV42515;  
 AC  
 XX  
 DT 05-OCT-1998 (first entry)  
 XX  
 XX PCR primer used to isolate part of the RELAX protein coding region.  
 DE  
 XX Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;  
 KW control; gene expression; transcriptional activator; targeting;  
 KW protein expression; central nervous system; CNS; treatment;  
 KW nervous system disorder; CIG235; PCR primer; ss.  
 XX  
 XX Synthetic.  
 OS  
 OS Rattus sp.  
 XX  
 XX W09827206-A2.  
 PN  
 XX 25-JUN-1998.  
 PD  
 XX 19-DEC-1997; 97WO-FR02368.  
 PF  
 XX 19-DEC-1996; 96FR-0015651.  
 PR  
 XX (RHON ) RHONE-POULENC RORER SA.  
 PA  
 XX Mallet J, Ravassard P, Icard-Liepkalns C;  
 PI  
 XX WPI; 1998-362775/31.  
 DR  
 XX Basic helix-loop-helix polypeptide and related nucleic acid - with  
 PT transcriptional activity, for targeting expression of genes to  
 PT central nervous system and treatment of nervous disease  
 PT  
 XX Example 2; Page 12; 28pp; French.  
 PS  
 XX PCR primers AAV42515-16 are used to isolate part of the DNA encoding  
 CC a basic helix-loop-helix (BHLH) type protein, designated RELAX (Rat  
 CC Embryonic Longitudinal Axis) protein. The PCR product is termed CIG235.  
 CC The protein is used to control and participate in gene expression,  
 CC by acting as transcriptional activator, strictly dependent on the

CC presence of an intact E box (CANWNG), particularly for targeting  
 CC expression of proteins to the central nervous system (CNS). The  
 CC nucleic acid sequence can be used to treat nervous system disorders,  
 CC and antisense sequences can be used to control mRNA transcription.

XX Sequence 25 BP; 5 A; 9 C; 6 G; 5 T; 0 other;

Query Match 1.7%; Score 25; DB 19; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.073;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 AACCTTAACCTCGCGTGGATGCGC 774

DB 1 AACCTTAACCTCGCGTGGATGCGC 25

RESULT 13

AAV27046

ID AAV27046 standard; cDNA; 738 BP.

XX AAV27046;

AC AAV27046;

XX 17-SEP-1998 (first entry)

DT Mouse neurogenin 1 gene.

DE Mouse neurogenin 1 gene.

XX ss; Mouse; neurogenin; expression vector; recombinant protein;

KW antibody; neurogenesis.

OS Mus sp.

XX Key

FT CDS

FT Location/Qualifiers

FT 1..735

FT /\*tag= a

FT /product= "Mouse neurogenin 1"

XX WO9813491-A2.

PN 02-APR-1998.

XX 24-SEP-1997; 97WO-US17048.

XX 17-SEP-1997; 97US-0932411.

PR 27-SEP-1996; 96US-0722570.

PR 12-NOV-1996; 96US-0030864.

PR 19-DEC-1996; 96US-0772009.

XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

PA Anderson DJ, Ma Q, Sommer L;

XX WPI; 1998-230702/20.

DR P-PSDB; AAW54944.

XX Mouse neurogenin1, useful in neurogenesis - and recombinant nucleic

PT acids and proteins derived from rat and Xenopus

XX Claim 5; Fig 4; 106pp; English.

PS The mouse neurogenin 1 is one of several neurogenin proteins discussed in

XX the present invention. The neurogenin nucleic acids can be expressed in

CC a host cell, transformed using an expression vector, to produce

CC recombinant proteins. The proteins and the antibodies raised against

CC the proteins are useful in the study of neurogenesis.

XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

QY Query Match 1.6%; Score 23; DB 19; Length 738;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754

|||||

DB 304 GAGCGCAACCGCATGCACACCT 326

RESULT 14

AAZ51977

ID AAZ51977 standard; DNA; 738 BP.

XX AAZ51977;

AC AAZ51977;

XX 04-JUL-2000 (first entry)

DT Murine neurogenin-1 (NGN1) nucleic acid sequence.

XX Neurogenin-1; NGN-1; non-neuronal cell; NNC; neurogenesis;

XX Phox2a protein; neuronal subtype-specific marker; growth factor;

KW neuronal differentiation; transplantation; neuronal dysfunction;

KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;

KW neuroprotective; nontropic; anticonvulsant; antiparkinsonian; vulnary;

KW cerebroprotective; immunosuppressant; antiinfectious; ss.

XX Mus sp.

OS Key

XX Location/Qualifiers

FT CDS

FT 1..735

FT /\*tag= a

FT /product= "Murine neurogenin-1 protein"

XX WO200009676-A2.

PN 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18525.

XX 14-AUG-1998; 98US-0096630.

XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

PA Anderson DJ, Lo L;

XX WPI; 2000-256250/22.

DR P-PSDB; AAY70566.

XX Inducing non-neuronal cells to differentiate into neurons and for

PT non-neuronal cells to express a neuronal subtype-specific marker,

PT comprising contacting the non-neuronal cells with a vector containing

PT neurogenin nucleic acid -

XX Claim 1; Fig 1C; 76pp; English.

PS The patent discloses a method for inducing non-neuronal cells (NNC) to

XX differentiate into neurons and for NNCs to express a neuronal subtype

CC -specific marker. Transformed host cells are used as sources of neuronal

CC and other growth factors; in culture for screening compounds that

CC modulate neural differentiation or as sources of recombinantly produced

CC neurogenins and Phox2a proteins for use in transplantation. The cells

CC also have a variety of in vivo uses, e.g. for transplantation at sites of

CC neuronal dysfunction e.g. patients with hearing or vision loss due to

CC optical or auditory nerve damage, brain or spinal cord injuries, and

CC neurodegenerative disorders e.g. Alzheimer's disease. The present

CC sequence encodes murine neurogenin-1 (NGN-1), a transcription factor.

CC NNCs differentiate into neurons through the recombinant expression of a

CC transcription factor that induces a core program of neurogenesis. Forced

CC expression of murine NGN1 can elicit expression of at least some neuronal

CC phenotypic markers even in NNCs. This can be used in autografting.

XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

QY Query Match 1.6%; Score 23; DB 21; Length 738;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754

|||||

Db 304 GAGCGCAACCGCATGCACACCT 326

Search completed: April 8, 2003, 19:05:41  
Job time : 407 secs

## RESULT 15

AAE27264  
ID AAE27264 standard; cDNA; 790 BP.  
XX  
AC AAE27264;  
XX  
XX 24-APR-2001 (first entry)  
XX  
DE Chicken atonal homologue ngn2/ath4a cDNA, SEQ ID NO:20.  
XX  
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;  
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
KW cellular proliferation; cerebellar granule neuron; gene therapy;  
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;  
KW transgenic animal; ss.  
XX  
OS Gallus gallus.  
XX  
XX WO2000073764-A2.  
XX  
XX 07-DEC-2000.  
XX  
XX 01-JUN-2000; 2000WO-US15410.  
XX  
XX 01-JUN-1999; 99US-0137060.  
XX  
XX 19-JAN-2000; 2000US-0176993.  
XX  
XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
XX  
XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
XX  
XX WPI; 2001-032190/04.  
XX  
XX P-PSDB; AAB60357.  
XX  
XX Therapeutic use of atonal-associated nucleic acids or amino acids, or  
XX any of its homologs or orthologs, for the treatment of e.g. deafness,  
XX osteoarthritis and abnormal cell proliferation -  
XX  
XX Disclosure; Page -: 142pp; English.  
XX  
XX The invention relates to the use of atonal-associated nucleic acid or  
XX amino acid sequence, or any of its homologs or orthologues as  
XX therapeutic agents for the treatment of deafness, partial hearing loss,  
XX vestibular effects due to damage or loss of inner hair cells,  
XX osteoarthritis and abnormal cell proliferation. The invention also  
XX encompasses methods of screening for compounds which affect the  
XX expression of an atonal-associated nucleic acid sequence in an animal,  
XX and a transgenic animal in which an allele of a native atonal-associated  
XX gene is replaced by a heterologous nucleic acid sequence, thus  
XX inactivating the atonal-associated allele. The nucleic acids or proteins  
XX may be used in a method of treating an animal for hearing impairment,  
XX joint disease, balance disorders, abnormal cell proliferation, or other  
XX disease related to loss of a functional atonal-associated nucleic acid or  
XX protein. They may particularly be used to treat an animal with a  
XX deficiency in cerebellar granule neurons or their precursors, and may  
XX also be used in promoting mechanoreceptive cell growth and generating  
XX hair cells. The present sequence represents an atonal-associated nucleic  
XX acid sequence referred to in the invention.  
XX  
XX Note: The present sequence is not shown in the specification, but  
XX was obtained from GenBank.

XX SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 790;

Best Local Similarity 100.0%; Pred. No. 0.59;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754

|||||

Db 374 GAGCGCAACCGCATGCACACCT 396

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 17:56:31 ; Search time 165 seconds  
(without alignments)  
7761.606 Million cell updates/sec

Title: US-09-595-947c-1

Perfect score: 1460

Sequence: 1 gcaggtacgagagagcag.....agagtacacacacacgtgt 1460

Scoring table: OLIGO.NUC

Gapop 60.0 , Gapext 60.0

Searched: 593429 seqs, 438583890 residues

Word size : 8

Total number of hits satisfying chosen parameters: 919487

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	6.3	645	9	US-10-004-7117-4
2	92	6.3	861	9	US-10-004-7117-24
3	92	6.3	1861	10	US-09-817-360-3
4	32	2.2	5340	10	US-09-817-360-1
5	23	1.6	738	7	US-08-722-570-13
6	23	1.6	790	9	US-10-004-7117-20
7	23	1.6	1385	9	US-10-004-7117-30
8	23	1.6	1412	9	US-10-004-7117-6
9	23	1.6	1412	9	US-10-004-7117-37
10	20	1.4	352	10	US-09-728-445-108
11	19	1.3	500	10	US-09-783-590-2157
12	19	1.3	1263	9	US-09-938-842A-1036
13	18	1.2	4797	10	US-09-751-797-25
14	18	1.2	20272	10	US-09-908-711-145
15	18	1.2	25619	10	US-09-908-711-143
16	18	1.2	25619	10	US-09-764-898-302
17	18	1.2	31728	9	US-10-114-170-64
18	18	1.2	402850	9	US-09-844-653-5
19	17	1.2	31	10	US-09-817-360-17

20	1.2	232	10	US-09-960-352-14228	Sequence 14228, A
21	1.2	285	10	US-09-960-352-6887	Sequence 6887, Ap
22	1.2	285	10	US-09-960-352-8778	Sequence 8778, Ap
c 23	1.2	307	10	US-09-783-590-12149	Sequence 12149, A
24	1.2	454	9	US-10-040-739-1094	Sequence 1094, Ap
c 25	1.2	454	10	US-09-864-761-257	Sequence 257, App
c 26	1.2	469	10	US-09-864-761-5374	Sequence 5374, Ap
c 27	1.2	471	10	US-09-864-761-11	Sequence 11, Appl
c 28	1.2	473	10	US-09-864-761-22148	Sequence 22148, A
c 29	1.2	474	10	US-09-864-761-353	Sequence 353, App
c 30	1.2	474	10	US-09-864-761-10932	Sequence 10932, A
c 31	1.2	487	10	US-09-864-761-3320	Sequence 3320, Ap
c 32	1.2	495	10	US-09-864-761-5673	Sequence 5673, Ap
c 33	1.2	499	10	US-09-783-590-4421	Sequence 4421, Ap
c 34	1.2	594	10	US-09-864-761-7938	Sequence 7938, Ap
c 35	1.2	756	10	US-09-910-943-147	Sequence 147, App
c 36	1.2	1074	9	US-10-004-7117-18	Sequence 18, Appl
c 37	1.2	1238	9	US-09-991-496-75	Sequence 75, Appl
c 38	1.2	1238	10	US-09-874-923-75	Sequence 75, Appl
c 39	1.2	1362	9	US-09-970-966-208	Sequence 208, App
c 40	1.2	1362	10	US-09-825-294-208	Sequence 208, App
c 41	1.2	1422	10	US-09-815-242-4334	Sequence 4334, Ap
c 42	1.2	1425	10	US-09-815-242-8558	Sequence 8558, Ap
c 43	1.2	1669	10	US-09-880-107-2159	Sequence 2159, Ap
c 44	1.2	1747	9	US-09-764-868-1334	Sequence 1334, Ap
c 45	1.2	1752	9	US-09-764-868-137	Sequence 137, App

#### ALIGNMENTS

#### RESULT 1

US-10-004-7117-4  
; Sequence 4, Application US/10004717  
; Publication No. US20020192665A1  
; GENERAL INFORMATION:  
; APPLICANT: ZOGHBI, HUDA Y.  
; APPLICANT: YANG, QI  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS.  
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
; FILE REFERENCE: P01899054  
; CURRENT APPLICATION NUMBER: US/10/004,717  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: 09/585,645  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/176,993  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: 60/137,060  
; PRIOR FILING DATE: 1999-06-01  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 645  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-004-7117-4

Query Match 6.3%; Score 92; DB 9; Length 645;  
Best Local Similarity 100.0%; Pred. No. 5.5e-39;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	762	GGCGTGGTGGCGTGGCGGTGTCCTCCGACCTTCCCGATGACGCCAACTTACAAG	821
Db	304	GGCGTGGTGGCGTGGCGGTGTCCTCCGACCTTCCCGATGACGCCAACTTACAAG	363
QY	822	ATCGAGACCTTGGCTTCCGCCCAACTATACAT	853
Db	364	ATCGAGACCTTGGCTTCCGCCCAACTATACAT	395

#### RESULT 2

US-10-004-7117-24

```

; Sequence 24, Application US/10004717
; Publication No. US20020192865A1
; GENERAL INFORMATION:
; APPLICANT: ZOGBBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P018990S4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-24

```

	Query Match	6.3%	Score 92	DB 9	Length 861
	Best Local Similarity	100.0%	Pred. No. 5.4e-39		
	Matches 92	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	762	GCGETGATGCGTGC	CGCGTGTCTCTGCCACCTTC	CCGGATGATCGCCAACTTACAAG	821
Db	463	GCGETGATGCGTGC	CGCGTGTCTCTGCCACCTTC	CCGGATGATCGCCAACTTACAAG	522
QY	822	ATCGAGACCGCTGC	CGCTTCGCCCACTACAT		853
Db	523	ATCGAGACCGCTGC	CGCTTCGCCCACTACAT		554

```

RESULT 3
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129C1P
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-817-360-3

```

			Query Match	6.3%	Score 92;	DB 10;	Length 1861;	
			Best Local Similarity	100.0%;	Pred. No.	5e-39;		
			Matches	92;	Conservative	0;	Mismatches	0; Indels
								0; Gaps
								0;
Qy	762	GCGTGTGATGGCTGC	CGGGTGTCCTGCCCAACTTC	CCGGATGACG	CCAAACTTTACAAG	821		
Dz	1396	GCGTGTGATGGCTGC	CGGGTGTCCTGCCCAACTTC	CCGGATGACG	CCAAACTTTACAAG	1455		
Qy	822	ATCAGAGCCCTGCGCTTC	CGCCCACAAC	TACAT	853			
Dz	1456	ATCAGAGCCCTGCGCTTC	CGCCCACAAC	TACAT	1487			

```

RESULT 4
US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 80/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-817-360-1

```

```

Query Match      2.2%; Score 32; DB 10; Length 5340;
Best Local Similarity 100.0%; Pred.No. 4.9e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 672 AAGACGAGTTGGCACTGAGCAAGCAGCGAGC 703
.      |||||
Db 3235 AAGACGAGTTGGCACTGAGCAAGCAGCGAGC 3266

```

RESULT 5  
US-08-722-570-13  
Sequence 13, Application US/08722570  
Publication No. US20030044887A1  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qiufu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-Sep-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown

```
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-722-570-13

Query Match      1.6%; Score 23; DB 7; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 304 GAGCGCAACCGCATGCACAACCT 326

RESULT 6
US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 790
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-20

Query Match      1.6%; Score 23; DB 9; Length 790;
Best Local Similarity 100.0%; Pred. No. 0.038;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 374 GAGCGCAACCGCATGCACAACCT 396

RESULT 7
US-10-004-717-30
; Sequence 30, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-30

Query Match      1.6%; Score 23; DB 9; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 742 GAGCGCAACCGCATGCACAACCT 764

RESULT 8
US-10-004-717-6
; Sequence 6, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-6

Query Match      1.6%; Score 23; DB 9; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754
|||||
Db 445 GAGCGCAACCGCATGCACAACCT 467

RESULT 9
US-10-004-717-37
; Sequence 37, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-37
```



US-10-004-717-37

Query Match 1.6%; Score 23; DB 9; Length 1412;  
Best Local Similarity 100.0%; Pred. No. 0.036;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754  
|||||  
DB 445 GAGCGCAACCGCATGCACAACT 467

RESULT 10

US-09-728-445-108/c  
; Sequence 108, Application US/09728445  
; Patent No. US20020102543A1  
; GENERAL INFORMATION:  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and  
; FILE REFERENCE: LEX-0102-USA  
; CURRENT APPLICATION NUMBER: US/09/728,445  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/168,358  
; PRIOR FILING DATE: 1999-12-01  
; NUMBER OF SEQ ID NOS: 891  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 108  
; LENGTH: 352  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-728-445-108

Query Match 1.4%; Score 20; DB 10; Length 352;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1010 CACAGCCTCATTTGGAGGACT 1029  
|||||  
DB 209 CACAGCCTCATTTGGAGGACT 190

RESULT 11

US-09-783-590-2157/c  
; Sequence 2157, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2157  
; LENGTH: 500  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (84)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (207)

; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (215)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (324)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (354)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (360)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (368)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (369)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (379)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (393)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (411)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (418)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (439)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (440)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (441)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (451)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-2157

Query Match 1.3%; Score 19; DB 10; Length 500;  
Best Local Similarity 100.0%; Pred. No. 5.4;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1319 CCACCTCAAACTCCCGCTC 1337  
|||||  
DB 23 CCACCTCAAACTCCCGCTC 5

RESULT 12

US-09-938-842A-1036/c  
; Sequence 1036, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAININ  
; FILE REFERENCE: SAME, AND METHODS OF USE  
; FILE REFERENCE: SCRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111

; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1036  
; LENGTH: 1263  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1036

Query Match 1.3%; Score 19; DB 9; Length 1263;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 658 GCGCAACAGGCCCAAGAG 676  
|||||  
DB 599 GCGCAACAGGCCCAAGAG 581

## RESULT 13

US-09-751-797-25  
; Sequence 25, Application US/09751797  
; Patent No. US20010024652A1  
; GENERAL INFORMATION:  
; APPLICANT: Dumoutier, Laure  
; APPLICANT: Louhed, Jamila  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Factors  
; FILE REFERENCE: LUD 5543.2  
; CURRENT APPLICATION NUMBER: US/09/751,797  
; CURRENT FILING DATE: 2000-12-29  
; PRIOR APPLICATION NUMBER: 09/419,568  
; PRIOR FILING DATE: 1999-10-18  
; PRIOR APPLICATION NUMBER: US09/178,973  
; PRIOR FILING DATE: 1998-10-26  
; NUMBER OF SEQ ID NOS: 29  
; SEQ ID NO 25  
; LENGTH: 4797  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-751-797-25

Query Match 1.2%; Score 18; DB 10; Length 4797;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1152 GTCAGAGCTGTCTGAAAT 1169  
|||||  
DB 3460 GTCAGAGCTGTCTGAAAT 3477

## RESULT 14

US-09-908-711-145/c  
; Sequence 145, Application US/09908711  
; Patent No. US20020045230A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA128  
; CURRENT APPLICATION NUMBER: US/09/908,711  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US01/01360  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,867  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01344  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01345  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,888  
; PRIOR FILING DATE: 2001-01-17

; PRIOR APPLICATION NUMBER: US01/01329  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,905  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01354  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01339  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,869  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01340  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,874  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01334  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,898  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01320  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,853  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01349  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,902  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01239  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,870  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01348  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,882  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01347  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,896  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01307  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,864  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01341  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,856  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01336  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,868  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01312  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 167  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 145  
; LENGTH: 20272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-711-145

Query Match 1.2%; Score 18; DB 10; Length 20272;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1153 TCAGAGCTGTCTGAAATG 1170

Db 6642 TCAGAGCTGTCTGAAATG 6625  
|||||

## RESULT 15

US-09-908-711-143/c  
; Sequence 143, Application US/09908711  
; Patent No. US20020045230A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA128  
; CURRENT APPLICATION NUMBER: US/09/908,711  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US01/01360  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,867  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01344  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01345  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,888  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01329  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,905  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01354  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01339  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,869  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01340  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,874  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01334  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,898  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01320  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,853  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01349  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,902  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01239  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,870  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01348  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,882  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01347  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,896  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01307  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,864  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01341  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,856

; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01336  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,868  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: US01/01312  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 167  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 143  
; LENGTH: 25619  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-711-143

Query Match 1.2%; Score 18; DB 10; Length 25619;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1153 TCAGAGCTGTCTGAAATG 1170  
|||||  
Db 1048 TCAGAGCTGTCTGAAATG 1031

Search completed: April 8, 2003, 21:14:29  
Job time : 291 secs

/db\_xref="taxon:9606"  
/clone="IMAGE:4110222"  
/clone\_lib="NIH\_MGC\_17"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 236 a 219 c 262 g 196 t  
ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 913;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 CAAGGGGTAGAAGGGGAGT 374  
|||||  
Db 853 CAAGGGGTAGAAGGGGAGT 873

RESULT 13  
LOCUS AU067624 953 bp mRNA linear EST 12-JUL-2000  
DEFINITION AU067624 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-3693  
5', mRNA sequence.  
ACCESSION AU067624  
VERSION AU067624.1 GI:4968361  
KEYWORDS EST.  
SOURCE mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 953)  
AUTHORS Hashimoto, K., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Toyoda, A.,  
Suzuki, Y., Sasaki, M. and Sugano, S.  
TITLE Isolation of full-length cDNA clones from a mouse brain cDNA  
library made by oligo-capping method  
JOURNAL Unpublished (1999)  
COMMENT Contact: Katsuyuki Hashimoto  
Division of Genetic Resources  
National Institute of Infectious Diseases  
23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
Email: khashi@nih.go.jp  
URL: http://www.nih.go.jp/yoken/genbank/  
LOCATION/Qualifiers  
FEATURES  
source  
1. .953  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="MNCB-3693"  
/clone\_lib="MNCB-3693"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="TOP10"  
/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA  
was primed with an oligo(dT) primer  
ATGTGGCCCTTTTCTTTTCTTTT; double-stranded cDNA was  
ligated to a DraIII adaptor [GTGGGCTACTGG], digested and  
cloned into distinct DraIII sites of the pME18S-FL3. XhoI  
sites just outside the DraIII sites can be used to isolate  
the cDNA insert. Size selection was performed to exclude  
fragments <1.5 kb. Library was constructed by Sugano et  
al. (University of Tokyo, Institute of Medical Science).  
Custom primer used for sequencing: 5' end primer  
[CTTCTGCTTAAAGCTCGG], 3' end primer  
[CGACTGCGAGTCGAGCAC]"  
BASE COUNT 186 a 271 c 276 g 156 t 64 others  
ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 953;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 CGGAGCGCACCGCATGTCAC 749  
|||||  
Db 334 CGGAGCGCACCGCATGTCAC 354

RESULT 14  
LOCUS BO686909 962 bp mRNA linear EST 15-JUL-2002  
DEFINITION BO686909 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6249314  
5', mRNA sequence.  
ACCESSION BO686909  
VERSION BO686909.1 GI:21812225  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 962)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Cloned by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM2390 row: 1 column: 03  
High quality sequence start: 3  
High quality sequence stop: 587.  
LOCATION/Qualifiers  
FEATURES  
source  
1. .962  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6249314"  
/clone\_lib="NIH\_MGC\_110"  
/tissue\_type="ductal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 236 a 273 c 259 g 194 t  
ORIGIN

Query Match 1.4%; Score 21; DB 14; Length 962;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 ACCCCCTTCATTTTTCCTCC 449  
|||||  
Db 811 ACCCCCTTCATTTTTCCTCC 831

RESULT 15  
LOCUS BM563662 1022 bp mRNA linear EST 20-FEB-2002  
DEFINITION BM563662 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5441798  
5', mRNA sequence.  
ACCESSION BM563662  
VERSION BM563662.1 GI:18810777

/db\_xref="taxon:9606"  
/clone="IMAGE:4110222"  
/clone\_lib="NIH\_MGC\_17"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pOTB7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 236 a 219 c 262 g 196 t  
ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 913;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 CAAGGGGTAGAAGGGGAGT 374  
|||||  
Db 853 CAAGGGGTAGAAGGGGAGT 873

RESULT 13  
LOCUS AU067624 953 bp mRNA linear EST 12-JUL-2000  
DEFINITION AU067624 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-3693  
5', mRNA sequence.  
ACCESSION AU067624  
VERSION AU067624.1 GI:4968361  
KEYWORDS EST.  
SOURCE mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 953)  
AUTHORS Hashimoto, K., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Toyoda, A.,  
Suzuki, Y., Sasaki, M. and Sugano, S.  
TITLE Isolation of full-length cDNA clones from a mouse brain cDNA  
library made by oligo-capping method  
JOURNAL Unpublished (1999)  
COMMENT Contact: Katsuyuki Hashimoto  
Division of Genetic Resources  
National Institute of Infectious Diseases  
23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan  
Email: khashi@nih.go.jp  
URL: http://www.nih.go.jp/yoken/genbank/  
LOCATION/Qualifiers  
FEATURES  
source  
1. .953  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="MNCB-3693"  
/clone\_lib="MNCB-3693"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="TOP10"  
/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA  
was primed with an oligo(dT) primer  
ATGTGGCCCTTTTCTTTTCTTTT; double-stranded cDNA was  
ligated to a DraIII adaptor [GTGGGCTACTGG], digested and  
cloned into distinct DraIII sites of the pME18S-FL3. XhoI  
sites just outside the DraIII sites can be used to isolate  
the cDNA insert. Size selection was performed to exclude  
fragments <1.5 kb. Library was constructed by Sugano et  
al. (University of Tokyo, Institute of Medical Science).  
Custom primer used for sequencing: 5' end primer  
[CTTCTGCTTAAAGCTCGG], 3' end primer  
[CGACTGCGAGTCGAGCAC]"  
BASE COUNT 186 a 271 c 276 g 156 t 64 others  
ORIGIN

Query Match 1.4%; Score 21; DB 9; Length 953;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 CGGAGCGCACCGCATGTCAC 749  
|||||  
Db 334 CGGAGCGCACCGCATGTCAC 354

RESULT 14  
LOCUS BO686909 962 bp mRNA linear EST 15-JUL-2002  
DEFINITION BO686909 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6249314  
5', mRNA sequence.  
ACCESSION BO686909  
VERSION BO686909.1 GI:21812225  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 962)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Cloned by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM2390 row: 1 column: 03  
High quality sequence start: 3  
High quality sequence stop: 587.  
LOCATION/Qualifiers  
FEATURES  
source  
1. .962  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6249314"  
/clone\_lib="NIH\_MGC\_110"  
/tissue\_type="ductal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 236 a 273 c 259 g 194 t  
ORIGIN

Query Match 1.4%; Score 21; DB 14; Length 962;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 ACCCCCTTCATTTTTCCTCC 449  
|||||  
Db 811 ACCCCCTTCATTTTTCCTCC 831

RESULT 15  
LOCUS BM563662 1022 bp mRNA linear EST 20-FEB-2002  
DEFINITION BM563662 NIH\_MGC\_98 Homo sapiens cDNA clone IMAGE:5441798  
5', mRNA sequence.  
ACCESSION BM563662  
VERSION BM563662.1 GI:18810777

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1022)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1916 row: k column: 15  
High quality sequence stop: 654.  
FEATURES  
Location/Qualifiers  
1..1022  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5441798"  
/clone\_lib="NIH\_MGC\_98"  
/tissue\_type="astrocytoma grade IV, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into ECORI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
BASE COUNT 177 a 324 c 319 g 202 t  
ORIGIN  
Query Match 1.4%; Score 21; DB 13; Length 1022;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1112 AACAGGCCCTGGCGGTGGGC 1132  
|||||  
Db 573 AACAGGCCCTGGCGGTGGGC 593  
Search completed: April 8, 2003, 21:09:45  
Job time : 2284 secs

pred. No. is the number of results predicted by chance to have a

301 GCAGCTCTGTCTCTTTTGAGCCCCGGAGTAAGTAACAT

pred. No. is the number of results predicted by chance to have a

301 GCAGCTCTCTGTTCTTTTGGCCGCGAGTAAGTAACTAGTAACATTTAGGAACCTCCAAAGG 360  
361 TAGAAGAGGGAGTGGTGGGCTACTCTAGTCCCGCGTGGAGTACCTCTAAGTCAGAG 420  
361 TAGAAGAGGGAGTGGTGGGCTACTCTAGTCCCGCGTGGAGTACCTCTAAGTCAGAG 420  
421 ACTGTCCACACCCCTCTCCATTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480  
421 ACTGTCCACACCCCTCTCCATTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480  
481 CGCCCAACATCAAGTGTCCCAAGAGACCCAGAACCTTTCCCGAGGCTCGGACCAGG 540  
481 CGCCCAACATCAAGTGTCCCAAGAGACCCAGAACCTTTCCCGAGGCTCGGACCAGG 540  
541 AGTGTCTCAGTTCCCAATTCACCCACCTAGCCCACTCTCTACCGAGGAGTCTCTCG 600  
541 AGTGTCTCAGTTCCCAATTCACCCACCTAGCCCACTCTCTACCGAGGAGTCTCTCG 600  
601 AAGCAGAAGCAGTGTACTCCGAGGACATCGAGGAAGCTCGTGGCGCGCGGAGGCG 660  
601 AAGCAGAAGCAGTGTACTCCGAGGACATCGAGGAAGCTCGTGGCGCGCGGAGGCG 660  
661 GCACAGGCGCCAAAGAGGAGTTGGCACTGAGCAAGCAGCAGAGCCCGGCCAAGAAGG 720  
661 GCACAGGCGCCAAAGAGGAGTTGGCACTGAGCAAGCAGCAGAGCCCGGCCAAGAAGG 720  
721 CCAACGACCGGAGCGCAACCGCATGCACAACTTAACCTCGGCTGGATCGCTCGCG 780  
721 CCAACGACCGGAGCGCAACCGCATGCACAACTTAACCTCGGCTGGATCGCTCGCG 780  
781 GTGCTCGCCCACTTCCCGGATGAGCCCAACTTACAAAGATCGAGACCTCGCGCTCG 840  
781 GTGCTCGCCCACTTCCCGGATGAGCCCAACTTACAAAGATCGAGACCTCGCGCTCG 840  
841 CCCACAACCTATTTGGGCACTGACTCAGACGCTGGCACTAGCGGACCACTCTTACG 900  
841 CCCACAACCTATTTGGGCACTGACTCAGACGCTGGCACTAGCGGACCACTCTTACG 900  
901 GCCCGAGCGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGG 960  
901 GCCCGAGCGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGG 960  
961 ACTGGGGCTCTATCTACTCCCGAGTTCCCAAGCTGTGAGCTGAGCCCTGAGCCCTCAT 1020  
961 ACTGGGGCTCTATCTACTCCCGAGTTCCCAAGCTGTGAGCTGAGCCCTGAGCCCTCAT 1020  
1021 TGGAGAGTTCCTGCGCTGAGGTCGCCAGCTCCCATCTCTGCTGCTCCCGGCGACCC 1080  
1021 TGGAGAGTTCCTGCGCTGAGGTCGCCAGCTCCCATCTCTGCTGCTCCCGGCGACCC 1080  
1081 TGGTGTCTCAGACTTCTGTGAAGGGCCCAACAGGCGCTGGGCGTGGCGCTGGCAG 1140  
1081 TGGTGTCTCAGACTTCTGTGAAGGGCCCAACAGGCGCTGGGCGTGGCGCTGGCAG 1140  
1141 AAGGAGGAGGATCAGAGTGTCTGAATGGAAGTGTGAGGCACTCGAGCATCTCGC 1200  
1141 AAGGAGGAGGATCAGAGTGTCTGAATGGAAGTGTGAGGCACTCGAGCATCTCGC 1200  
1201 CCTTCTGCTTCTTCTAGTCAAGTCCCTGATTTAACAGGATTCGCAGTTCCTTCT 1260  
1201 CCTTCTGCTTCTTCTAGTCAAGTCCCTGATTTAACAGGATTCGCAGTTCCTTCT 1260  
1261 GCTGTGCTGCAACAAGGACATTTGAGGCTGATCTCTCTTAACCTCTCTAGTGTGGC 1320  
1261 GCTGTGCTGCAACAAGGACATTTGAGGCTGATCTCTCTTAACCTCTCTAGTGTGGC 1320  
1321 ACCTCAAACTCCCGTCCCAAGCAGAGGAGCGGTAGCACTAAATGTTGGAGACTCC 1380  
1321 ACCTCAAACTCCCGTCCCAAGCAGAGGAGCGGTAGCACTAAATGTTGGAGACTCC 1380  
1381 ATACTTCTGTTGACTCCCGCTCTTCAATCTCGGCGCTTCAACACCGCTTCTCC 1440  
1381 ATACTTCTGTTGACTCCCGCTCTTCAATCTCGGCGCTTCAACACCGCTTCTCC 1440

QY 1441 AGAGTGACCTAATCCAGTGT 1460  
Db 1441 AGAGTGACCTAATCCAGTGT 1460  
RESULT 2  
RNRELAXT  
LOCUS R.norvegicus mRNA for transcriptional regulator, Relax.  
DEFINITION Y10619  
ACCESSION Y10619  
VERSION GI:2072737  
KEYWORDS Relax; transcriptional regulator.  
SOURCE Rattus norvegicus.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 1491)  
AUTHORS Ravassard, P., Chetail, F., Mallet, J. and Icard-Liepkalos, C.  
TITLE Relax, a novel rat BHLH transcriptional regulator transiently  
expressed in the ventricular proliferating zone of the developing  
central nervous system  
JOURNAL J. Neurosci. Res. 48 (2), 146-158 (1997)  
MEDLINE 97276390  
PUBMED 9130143  
REFERENCE 2 (bases 1 to 1491)  
AUTHORS Ravassard, P.  
TITLE Direct Submission  
JOURNAL Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERV1,  
Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013  
Paris, FRANCE  
FEATURES  
source  
1..1491  
/organism="Rattus norvegicus"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
/clone\_lib="RT-PCR"  
459..1103  
/function="transcriptional regulator"  
/codon\_start=1  
/product="Relax"  
/protein\_id="CAA71630.1"  
/db\_xref="GI:2072738"  
/db\_xref="SPTREMBL:O08718"  
/translation="MAPHPDAPTIQVQETQOPFGASDHEVLSSNPPSPFTLVPR  
DCSEAAGDCRGTSKKLRAGGRNRKSELASQKRRRRKANDREKRRHNLNSA  
LDLUNGVLPTPDDAKLTKIETLFAHNYIWAQTLLRIADHSFYGPPEPPVPGELGS  
PGGSSGDWGSIIYSPVQAGSLSPASLEFFPGLOVSPSPSCLLPGLTIVFSDL"  
BASE COUNT 307 a 487 c 413 g 284 t  
ORIGIN  
Query Match 100.0%; Score 1460; DB 10; Length 1491;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCAGGTAGCGAGGAGGAGTCCCTGGGCGCCGCTGATTGGCCGCGGACAGGCA 60  
Db 1 GCAGGTAGCGAGGAGGAGTCCCTGGGCGCCGCTGATTGGCCGCGGACAGGCA 60  
QY 61 GCAGCCCGGCGAGGACGCTCTGGTCCGGGCGAGAGCAGATAAGCGTCCAGGGGACACA 120  
Db 61 GCAGCCCGGCGAGGACGCTCTGGTCCGGGCGAGAGCAGATAAGCGTCCAGGGGACACA 120  
QY 121 CGATTAGAGCTCAGAGTCCCTCTGGGTCTCACCACCTCCAGAGGCGGAGACCCCT 180  
Db 121 CGATTAGAGCTCAGAGTCCCTCTGGGTCTCACCACCTCCAGAGGCGGAGACCCCT 180  
QY 181 CGAGCTCTTTCGTGCTCCAGAGCAATTTACTCCAGGCGAGGCGCTCGAGCTCAG 240  
Db 181 CGAGCTCTTTCGTGCTCCAGAGCAATTTACTCCAGGCGGAGGCGCTCGAGCTCAG 240  
QY 241 CAAAACCTCGAAGGAGGAGGAGGTTACGCTATCCACCGCTGCTTACTGTGACCTGACACCC 300

```
|||||
Db 241 CAAAACCTCGAAGCAGCAGAGGGGTTGAGTATCCACCGCTGCTGACTCTGACCAACC 300
QY 301 GCAGCTCTCTGTTCTTTTGGAGCCCGAGTAAGTAACTTTAGGAACCTTCCAAAGG 360
Db 301 GCAGCTCTCTGTTCTTTTGGAGCCCGAGTAAGTAACTTTAGGAACCTTCCAAAGG 360
QY 361 TAGAAGAGGGAGTGGTGGGCTACTAGTCCCGCTGGAGTGAACCTTAAGTCAGAG 420
Db 361 TAGAAGAGGGAGTGGTGGGCTACTAGTCCCGCTGGAGTGAACCTTAAGTCAGAG 420
QY 421 ACTGTACACACCCCTTCCATTTTCCCAACCTCAGATGCGCTCATCCTTTGGATG 480
Db 421 ACTGTACACACCCCTTCCATTTTCCCAACCTCAGATGCGCTCATCCTTTGGATG 480
QY 481 CGCCACCAATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGAGCCTCGGACCAG 540
Db 481 CGCCACCAATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCCCGAGCCTCGGACCAG 540
QY 541 AAGTGTCTAGTTCCAAATTCACCCACCTAGCCCACTCTCGTACCGAGGACTGCTCG 600
Db 541 AAGTGTCTAGTTCCAAATTCACCCACCTAGCCCACTCTCGTACCGAGGACTGCTCG 600
QY 601 AAGCAGAAGCAGTGAAGTGGGAGCATCGAGGAAGTCCGTGCGCGCGCGGAGGGC 660
Db 601 AAGCAGAAGCAGTGAAGTGGGAGCATCGAGGAAGTCCGTGCGCGCGCGGAGGGC 660
QY 661 GCAACAGGCCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCAGCAAGCCCGCGCAAGAAG 720
Db 661 GCAACAGGCCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCAGCAAGCCCGCGCAAGAAG 720
QY 721 CCAACGACCGGGAGCGCAACCGCATGCACAACTTAACCTCCGCTGGATGCGCTGGCG 780
Db 721 CCAACGACCGGGAGCGCAACCGCATGCACAACTTAACCTCCGCTGGATGCGCTGGCG 780
QY 781 GTGTCTGCCCCACTTCCCGGATGAGCCAACTTACAAGATCGAGACCTCGCTTCG 840
Db 781 GTGTCTGCCCCACTTCCCGGATGAGCCAACTTACAAGATCGAGACCTCGCTTCG 840
QY 841 CCCACAACACTACATTTGGGCACTGACTCAGACGCTGGCGATAGCGGACCAAGCTTCTACG 900
Db 841 CCCACAACACTACATTTGGGCACTGACTCAGACGCTGGCGATAGCGGACCAAGCTTCTACG 900
QY 901 GCGCCGAGCCCTGTGCGCTGTGGGAGCTGGAGCCCGGGAGGGGCTCCAGCGGG 960
Db 901 GCGCCGAGCCCTGTGCGCTGTGGGAGCTGGAGCCCGGGAGGGGCTCCAGCGGG 960
QY 961 ACTGGGGCTCTATCTACTCCCAAGTTTCCCAAGCTGGTAGCTGAGCCCAAGCCCTCAT 1020
Db 961 ACTGGGGCTCTATCTACTCCCAAGTTTCCCAAGCTGGTAGCTGAGCCCAAGCCCTCAT 1020
QY 1021 TGGAGGAGTTCCCTGGCTGCAAGTGGCCAGCTCCCATCCTGTCTCTCCCGGGCACCC 1080
Db 1021 TGGAGGAGTTCCCTGGCTGCAAGTGGCCAGCTCCCATCCTGTCTCTCCCGGGCACCC 1080
QY 1081 TGGTGTCTCAGACTTCTGTGAAGGCCCAACAGGCCCTGGCGGTGGCGCTGGCAG 1140
Db 1081 TGGTGTCTCAGACTTCTGTGAAGGCCCAACAGGCCCTGGCGGTGGCGCTGGCAG 1140
QY 1141 AAAGGGAGGAGTCAGAGCTGTCTGAATGGAAGTGTGGAGGACTCGAGCATCTCGC 1200
Db 1141 AAAGGGAGGAGTCAGAGCTGTCTGAATGGAAGTGTGGAGGACTCGAGCATCTCGC 1200
QY 1201 CCTTCTGGCTTTTCATTTAGTCAGTCCCTGATTTAACAGGATTCGCACAGTTCCTTGCT 1260
Db 1201 CCTTCTGGCTTTTCATTTAGTCAGTCCCTGATTTAACAGGATTCGCACAGTTCCTTGCT 1260
QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTTAACCCCTCCTCAGTGTGGCC 1320
Db 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTTAACCCCTCCTCAGTGTGGCC 1320
QY 1321 ACCTCAAACTCCCGCTCCAGCAGGAGGAGCCGTAGCACTAAATAGTTGGAGACTCCC 1380
|||||
```

Search completed: April 8, 2003, 17:16:12  
Job time : 4065 secs



ID AAV42512 standard; cDNA; 1491 BP.

Db 241 CAAACTTCGAAGCAGCAGAGGGGTTTCAGCTATCCACCGCTGCTGACTGTGACCTGACCAACC 300  
QY 301 GCAGCTCTCTGTCTTTTGGAGCCGGAGTAATAAGTAACATTTAGGAACCTCCAAAGGG 360  
Db 301 GCAGCTCTCTGTCTTTTGGAGCCGGAGTAATAAGTAACATTTAGGAACCTCCAAAGGG 360  
QY 361 TAGAAGAGGGAGTGGGTGGCGTACTCTAGTCCCGGTGGAGTGACCTCTAAGTCAGAG 420  
Db 361 TAGAAGAGGGAGTGGGTGGCGTACTCTAGTCCCGGTGGAGTGACCTCTAAGTCAGAG 420  
QY 421 ACTGTACACCCCTTCCATTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480  
Db 421 ACTGTACACCCCTTCCATTTTCCCAACCTCAGATGGCGCTCATCCCTTGGATG 480  
QY 481 CGCCACCATCCAGTGTCCCAAGAGACCAGCAACCTTTCCCGGAGCTCGGACCAAG 540  
Db 481 CGCCACCATCCAGTGTCCCAAGAGACCAGCAACCTTTCCCGGAGCTCGGACCAAG 540  
QY 541 AAGTGCTCAGTTCATTTCCACCCCTAGCCCTAGCCCACTCTCTAGCAGGAGTGTCTCG 600  
Db 541 AAGTGCTCAGTTCATTTCCACCCCTAGCCCTAGCCCACTCTCTAGCAGGAGTGTCTCG 600  
QY 601 AAGCAGAGCAGTGTACTGCCGAGGACATCGAGGAGTCCGTGCGCGCGGAGGGC 660  
Db 601 AAGCAGAGCAGTGTACTGCCGAGGACATCGAGGAGTCCGTGCGCGCGGAGGGC 660  
QY 661 GCAACAGGCCCAAGAGCGAGTGGCACTGAGCAAGCAGCAGCAAGCCGCGCGCAAGAAG 720  
Db 661 GCAACAGGCCCAAGAGCGAGTGGCACTGAGCAAGCAGCAGCAAGCCGCGCGCAAGAAG 720  
QY 721 CCAACGAGCCGAGCGCAACCCGATGACCAACCTTAACCTCCGCGTGGATGCGCTGCGG 780  
Db 721 CCAACGAGCCGAGCGCAACCCGATGACCAACCTTAACCTCCGCGTGGATGCGCTGCGG 780  
QY 781 GTGTCTGCCACCTTCCGAGTACGCGCAAACTTACAAAGATCGAGACCTTGCCTTCG 840  
Db 781 GTGTCTGCCACCTTCCGAGTACGCGCAAACTTACAAAGATCGAGACCTTGCCTTCG 840  
QY 841 CCCACAACCTACATTTGGGCACTGACTCAGACGCTCGCATAGCGGACCAAGCTTCTACG 900  
Db 841 CCCACAACCTACATTTGGGCACTGACTCAGACGCTCGCATAGCGGACCAAGCTTCTACG 900  
QY 901 GCGCCGAGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGCG 960  
Db 901 GCGCCGAGCCCTGTGCGCTGTGGGAGCTGGGAAGCCCGGAGGGGCTCCAGCGGCG 960  
QY 961 ACTGGGCTCTATCTACTCCCAAGTTCCCAAGCTGTAGCTGAGCCCAAGCTCAT 1020  
Db 961 ACTGGGCTCTATCTACTCCCAAGTTCCCAAGCTGTAGCTGAGCCCAAGCTCAT 1020  
QY 1021 TGGAGGAGTTCCTGGCTGAGTGCAGCTCCGAGCTCCCATCTGTCTCTCCCGGACCC 1080  
Db 1021 TGGAGGAGTTCCTGGCTGAGTGCAGCTCCGAGCTCCCATCTGTCTCTCCCGGACCC 1080  
QY 1081 TGGTGTCTCAGACTCTTGTGAAGGCCCAACAGGCGCTGGCGGTGGCGCTGGCAG 1140  
Db 1081 TGGTGTCTCAGACTCTTGTGAAGGCCCAACAGGCGCTGGCGGTGGCGCTGGCAG 1140  
QY 1141 AAAGGGAGGAGTCAAGCTGTCTGAATGAAGGTAGTGAAGGAGTCAAGGAGTCAAGCTTCGC 1200  
Db 1141 AAAGGGAGGAGTCAAGCTGTCTGAATGAAGGTAGTGAAGGAGTCAAGGAGTCAAGCTTCGC 1200  
QY 1201 CCTTCTGGCTTTCATAGTCAAGTCCCTGATTTAACAGGATTCGACAGTTCTCTTGT 1260  
Db 1201 CCTTCTGGCTTTCATAGTCAAGTCCCTGATTTAACAGGATTCGACAGTTCTCTTGT 1260  
QY 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAACCTCTCAGTGTGGCC 1320  
Db 1261 GCTGTGCTGCACAAAGGACATTCAGGCTGATCTCTTAACCTCTCAGTGTGGCC 1320  
QY 1321 ACCTCAACTCCGCTCCCAAGCAGAGGAGCGGTAGCACTAAATAGTTGGGAGACTCCC 1380  
Db 1321 ACCTCAACTCCGCTCCCAAGCAGAGGAGCGGTAGCACTAAATAGTTGGGAGACTCCC 1380

QY 1381 ATACTTCTGTGACTCGGCCCTCTTTCAAATCTCGGGCCTCCAAACACCGCTTTCTCC 1440  
Db 1381 ATACTTCTGTGACTCGGCCCTCTTTCAAATCTCGGGCCTCCAAACACCGCTTTCTCC 1440  
QY 1441 AGAGTGACCTAATCCAGTGT 1460  
Db 1441 AGAGTGACCTAATCCAGTGT 1460

Search completed: April 8, 2003, 11:37:01  
Job time : 374 secs

GenCore version 5.1.4.p5.4578  
Copyright(c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 11:23:35 ; Search time 122 Seconds  
(without alignments)  
10497.254 Million cell updates/sec

Title: US-09-595-947C-1  
Perfect score: 1460  
Sequence: 1 gcaggtagcagagagcag.....agagtacctaataatccagtg 1460

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

No matches found

Search completed: April 8, 2003, 18:58:39  
Job time : 124 secs

GenCore version 5.1.4\_p5\_4578  
Copyright(c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 06:37:40 ; Search time 2258 Seconds  
(without alignments)  
10471.838 Million cell updates/sec

Title: US-09-595-947C-1  
Perfect score: 1460  
Sequence: 1 gcaggtacgagagagcag.....agagtacctaataccagtg 1460

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 80%  
Maximum Match 100%  
Listing first 45 summaries

- Database : EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_estl:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vrt:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_Other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

No matches found

Search completed: April 8, 2003, 17:56:21  
Job time : 2258 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2003, 17:16:22 ; Search time 2260 Seconds  
(without alignments)  
10462.571 Million cell updates/sec

Title: US-09-595-947C-1  
Perfect score: 1460  
Sequence: 1 gcaggtagcagagagcag.....agagtgcacctaaccagtgt 1460

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 8

Total number of hits satisfying chosen parameters: 31743302

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*\*
  - 2: em\_esthum:\*\*
  - 3: em\_estin:\*\*
  - 4: em\_estmu:\*\*
  - 5: em\_estov:\*\*
  - 6: em\_estpl:\*\*
  - 7: em\_estro:\*\*
  - 8: em\_htc:\*\*
  - 9: gb\_est1:\*\*
  - 10: gb\_est2:\*\*
  - 11: gb\_htc:\*\*
  - 12: gb\_est3:\*\*
  - 13: gb\_est4:\*\*
  - 14: gb\_est5:\*\*
  - 15: em\_estfun:\*\*
  - 16: em\_estom:\*\*
  - 17: gb\_gss:\*\*
  - 18: em\_gss\_hum:\*\*
  - 19: em\_gss\_inv:\*\*
  - 20: em\_gss\_pln:\*\*
  - 21: em\_gss\_vrt:\*\*
  - 22: em\_gss\_fun:\*\*
  - 23: em\_gss\_man:\*\*
  - 24: em\_gss\_mus:\*\*
  - 25: em\_gss\_other:\*\*
  - 26: em\_gss\_pro:\*\*
  - 27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	6.3	593	17	AZ296526
2	60	4.1	1540	11	AK008017
3	23	1.6	600	12	BC808248
4	23	1.6	804	9	AL540071
5	23	1.6	814	14	BQ178789
6	23	1.6	1269	12	BG854922

7	21	1.4	319	12	BE936551
8	21	1.4	464	17	AQ753599
9	21	1.4	536	17	AZ506899
10	21	1.4	718	10	BE263765
11	21	1.4	829	13	BE191012
12	21	1.4	913	12	BF204175
13	21	1.4	953	9	AQ067624
14	21	1.4	962	14	BQ686909
15	21	1.4	1022	13	BM563662
16	21	1.4	1039	12	BE780690
17	21	1.4	1042	12	BG419220
18	20	1.4	241	9	AA113743
19	20	1.4	351	14	T70947
20	20	1.4	397	13	BI445873
21	20	1.4	444	10	AV944801
22	20	1.4	476	13	BM253695
23	20	1.4	550	10	AV939670
24	20	1.4	971	17	CNS01YSD
25	20	1.4	971	17	CNS02YDS
26	20	1.4	1075	14	BM810917
27	19	1.3	201	10	BB071548
28	19	1.3	234	10	BB269747
29	19	1.3	260	10	BE068280
30	19	1.3	301	10	BB305111
31	19	1.3	366	10	AW426722
32	19	1.3	370	10	AW481542
33	19	1.3	468	17	AZ286451
34	19	1.3	496	12	BF191528
35	19	1.3	507	12	BF408307
36	19	1.3	529	17	AZ830930
37	19	1.3	533	17	AZ985482
38	19	1.3	541	9	AU057600
39	19	1.3	552	13	BM030263
40	19	1.3	555	10	BE600721
41	19	1.3	564	13	BM088009
42	19	1.3	570	17	AQ594075
43	19	1.3	577	10	AW672224
44	19	1.3	579	10	AV597322
45	19	1.3	585	17	AQ631359

ALIGNMENTS

RESULT 1  
AZ296526/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AZ296526 593 bp DNA linear GSS 27-JUL-2000  
RPCI-23-160G18-TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18  
DNA sequence.  
AZ296526 GI:9538311  
GSS.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 593)  
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akintret  
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23  
Unpublished (1999)  
Other GSSs: RPCI-23-160G18.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from



```

/strain="C57BL/6J"
/db_xref="FANTOM.DB:2010001M19"
/db_xref="MGI:1907403"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/clone_lib="PIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..1540
/gene="Atoh5"
241..885
/gene="Atoh5"
/notes="atonal homolog 5 (Drosophila)"
data source:MGI, source key:MGI:893591, evidence:ISS
putative"
/codon_start=1
/protein_id="BAB25411.1"
/db_xref="GI:12841942"
/db_xref="MGI:893591"
/translation="MAPHPDALTIOVSPETQOPFCASDHEVLSNSTPSPHTLIPR
DCSEAEVGCDCGSRKLRARRGRNRPKSELAKRRSRKRRKANDRRNRNMNLISA
LDALRGVLPFPDPAKLTKVETLRFPAHYIWAQTIRIADHSLYGPEPPVPCGELGS
PGGSGNDGSIYSPVSGAGNLSPTASLEEFGLQVPSYLLPGALVFSDFL"
BASE COUNT 306 a 485 c 397 g 352 t
ORIGIN

Query Match 4.1%; Score 60; DB 11; Length 1540;
Best Local Similarity 100.0%; Pred. No. 4.8e-20;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGTGTGATGCGTGGCGGTGTCTGCCACCTCCCGGATGAGCGCAACTACAAAG 821
|||||
DB 544 GCGTGTGATGCGTGGCGGTGTCTGCCACCTCCCGGATGAGCGCAACTACAAAG 603
|||||

RESULT 3
BG080248 600 bp mRNA linear EST 20-DEC-2001
LOCUS 2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
DEFINITION mRNA sequence.
ACCESSION BG080248
VERSION BG080248.1 GI:17955225
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
Nucleic Acids Res. 29 (24), 4983-4993 (2001)
21671825
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
Location/Qualifiers
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dt. RNA isolation: cytoplasmic RNA preps
(Mannatis); Cloning technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TACGTCCACTGAATTCGAGTG--->. Other

BASE COUNT 154 a 270 c 278 g 99 t 3 others
ORIGIN

Query Match 1.6%; Score 23; DB 9; Length 804;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CTGCGCTTCGCCACCACTACAT 853
|||||
DB 770 CTGCGCTTCGCCACCACTACAT 792
|||||

RESULT 5
BQ178789 814 bp mRNA linear EST 30-APR-2002
LOCUS BQ178789
DEFINITION UI-M-EV0-bwt-k-07-0-UI.r1 NIH_BMAP-EV0 Mus musculus cDNA clone
IMAGE:5701758 5', mRNA sequence.
ACCESSION BQ178789
VERSION BQ178789.1 GI:20354281

information regarding entire library may be found at
http://pga.swmed.edu/Data/libraries/microarray_cdna_librar
ies.htm."
BASE COUNT 94 a 238 c 161 g 107 t
ORIGIN

Query Match 1.6%; Score 23; DB 12; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
|||||
DB 4 GAGCGCAACCGCATGCACACCT 26
|||||

RESULT 4
AL540071 804 bp mRNA linear EST 16-FEB-2001
LOCUS AL540071 LTI_FL013_FBrn1 Homo sapiens cDNA clone CS0DF035YD19 5
DEFINITION prime, mRNA sequence.
ACCESSION AL540071
VERSION AL540071.1 GI:12869886
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 804)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF035YD19"
/clone_lib="LTI_FL013_FBrn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH10B"
/notes="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 154 a 270 c 278 g 99 t 3 others
ORIGIN

Query Match 1.6%; Score 23; DB 9; Length 804;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 831 CTGCGCTTCGCCACCACTACAT 853
|||||
DB 770 CTGCGCTTCGCCACCACTACAT 792
|||||

RESULT 5
BQ178789 814 bp mRNA linear EST 30-APR-2002
LOCUS BQ178789
DEFINITION UI-M-EV0-bwt-k-07-0-UI.r1 NIH_BMAP-EV0 Mus musculus cDNA clone
IMAGE:5701758 5', mRNA sequence.
ACCESSION BQ178789
VERSION BQ178789.1 GI:20354281

```

```

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

FEATURES source
Seq primer: pYX-5.
Location/Qualifiers
1. .814
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5701758"
/clone_lib="NIH_BMAP_EV0"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is GTGCGTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP); 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 151 a 268 c 291 g 102 t 2 others
ORIGIN

Query Match 1.6%; Score 23; DB 14; Length 814;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
LOCUS |||||||||||||||||||
DB 663 GAGCGCAACCGCATGCACACCT 685

RESULT 6
BG854922 1269 bp mRNA linear EST 29-MAY-2001
DEFINITION 1024041C06.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BG854922.1 GI:14236106
VERSION BG854922
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
1 (bases 1 to 1269)
Chlamydomonadaceae; Chlamydomonas.
REFERENCE Grossman.A., Davies.J., Federspiel.N., Harris.E., Lefebvre.P.,
AUTHORS McDermott.J.P., Silflow,C., Stern,D. and Surzycki,R.

ANALYSES OF THE CHLAMYDOMONAS REINHARDTII GENOME: A MODEL,
UNICELLULAR SYSTEM FOR ANALYZING GENE FUNCTION AND REGULATION IN
VASCULAR PLANTS; PROJECT PHASE 2
UNPUBLISHED (2000)
CONTACT: CHARLES HAUSER
DCMB BOX 91000
DUKE UNIVERSITY
DURHAM, NC 27708-1000
TEL: 919 613 8159
FAX: 919 613 8177
EMAIL: CHAUSER@DUKE.EDU.
Location/Qualifiers
1. .1269
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP (acetate-containing) medium in the
light, TAP medium in the dark, HS (minimal) medium in
ambient levels of CO2 and HS medium bubbled with 5% CO2.
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with EXAssist (Stratagene)
phase. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 558 a 231 c 444 g 12 t 24 others
ORIGIN

Query Match 1.6%; Score 23; DB 12; Length 1269;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 653 CGGAGGGCGCACAGCCCAAGA 675
LOCUS |||||||||||||||||||
DB 1244 CGGAGGGCGCACAGCCCAAGA 1266

RESULT 7
BE936551 319 bp mRNA linear EST 02-OCT-2000
DEFINITION RC5-NT0053-310800-024-E01 NT0053 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE936551
VERSION BE936551.1 GI:10462627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 319)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

```



This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-rc5-NT0053-310  
 800-024-E01st3-2000-08-31st4-1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 319.

## FEATURES

Source

1. .319  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NT0053"  
 /dev\_stage="Adult"

/note="Organ: nervous\_tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 53 a 87 c 106 g 73 t

## ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 319;  
 Best Local Similarity 100.0%; Pred. NO. 15;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGC 1132

|||||

Db 280 AACAGGCCCTGGCGGTGGC 300

## RESULT 8

AQ753599

LOCUS

DEFINITION HS\_2117\_A2\_H03\_MR\_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2117 Col-6 Row-O, DNA sequence.

ACCESSION

AQ753599

VERSION

AQ753599.1 GI:5578650

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 464)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.

TITLE

Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome

JOURNAL

MEDLINE

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887

Email: jwallaceu.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: http://www.htsc.washington.edu

Plate: 2117 row: 0 column: 6

Seq primer: M13 Reverse

Class: BAC ends

High quality sequence stop: 464.

Location/Qualifiers

1. .464

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_plate="2117 Col-6 Row-O"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in  
 E-Coli DH10B"

BASE COUNT 125 a 114 c 101 g 123 t

## ORIGIN

Query Match 1.4%; Score 21; DB 17; Length 464;

Best Local Similarity 100.0%; Pred. NO. 16;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 492 CAAGTGTCCCAAGAGACCCAG 512

|||||

Db 106 CAAGTGTCCCAAGAGACCCAG 126

## RESULT 9

AZ506899

LOCUS

DEFINITION

1M0348G13F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0348G13 F, DNA sequence.

ACCESSION

AZ506899

VERSION

AZ506899.1 GI:10688215

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 536)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weiss,R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL

COMMENT

Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: dunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0348 row: G column: 13  
 Seq primer: CGTGTAAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 536.  
 Location/Qualifiers

1. .536

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC1M0348G13"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

/note="vector: PWD42uv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a  
 10.5 kb range using preparative agarose  
 electrophoresis. Vector DNA was prepared  
 of PWD42 (g14732114|gb|AF129072.1)  
 with adaptors complementary to the  
 purified. The sheared, adaptor mo  
 adapted vector DNA, and transforme  
 chemically-competent E. coli XL10-Go  
 and selected for ampicillin resistance;

```

BASE COUNT      156 a      119 c      81 g      180 t
ORIGIN

Query Match      1.4%; Score 21; DB 17; Length 536;
Best Local Similarity 100.0%; Pred. NO. 17;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 TGCCTGTGGGAGCTGGGAA 936
|||||
Db 406 TGCCTGTGGGAGCTGGGAA 426

RESULT 10
BE263765      718 bp      mRNA      linear      EST 13-JUL-2000
LOCUS 6011941122F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3537940 5',
DEFINITION mRNA sequence.
ACCESSION BE263765
VERSION BE263765.1 GI:9137311
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 718)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Plate: LNCM222 row: 1 column: 05
High quality sequence stop: 649.
Location/Qualifiers
1..718
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/lab_host="MGC3"
/Note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 115 a 195 c 260 g 148 t

ORIGIN

Query Match 1.4%; Score 21; DB 10; Length 718;  
Best Local Similarity 100.0%; Pred. NO. 18;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGGC 1132  
|||||  
Db 536 AACAGGCCCTGGCGGTGGGC 556

RESULT 11  
BI910102 829 bp mRNA linear EST 16-OCT-2001  
LOCUS 603067946F1 NIH\_MGC\_118 Homo sapiens cDNA clone IMAGE:5217154 5',  
DEFINITION mRNA sequence.  
ACCESSION BI910102  
VERSION BI910102.1 GI:16173443  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 829)  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11545 row: k column: 11  
High quality sequence stop: 828.  
Location/Qualifiers  
1..829  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_118"  
/tissue\_type="leukocyte"  
/lab\_host="DH108"  
/Note="Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV  
(destroyed); RNA source leukocytes from anonymous pool of  
non-activated adult donors. Library is oligo-dT primed  
and directionally cloned (EcoRV site is destroyed upon  
cloning). Average insert size 1.7 kb, insert size range  
1.2-3.3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 027. Note:  
this is a NIH\_MGC Library."

BASE COUNT 146 a 224 c 293 g 166 t

ORIGIN

Query Match 1.4%; Score 21; DB 13; Length 829;  
Best Local Similarity 100.0%; Pred. NO. 18;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCCCTGGCGGTGGGC 1132  
|||||  
Db 550 AACAGGCCCTGGCGGTGGGC 570

RESULT 12  
BF204175 913 bp mRNA linear EST 06-NOV-2000  
LOCUS 601867625F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4110222 5',  
DEFINITION mRNA sequence.  
ACCESSION BF204175  
VERSION BF204175.1 GI:11097761  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 913)  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LNCM996 row: a column: 07  
High quality sequence stop: 714.  
Location/Qualifiers  
1..913  
/organism="Homo sapiens"